

Kjetil Holm

Barnehage og leseferdigheter

Analyse av betydningen av å ha gått i barnehage (kindergarden) på elevprestasjoner.

Mai 2020

NTNU

Norges teknisk-naturvitenskapelige universitet.

Fakultet for økonomi

Institutt for samfunnsøkonomi

Bacheloroppgave

2020



Kjetil Holm

Barnehage og leseferdigheter

Analyse av betydningen av å ha gått i barnehage
(kindergarden) på elevprestasjoner.

Bacheloroppgave
Mai 2020

NTNU

Norges teknisk-naturvitenskapelige universitet.

Fakultet for økonomi

Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Contents

1 Innledning	1
2 Tidligere studier	1
3 Datamaterialet	3
4 Empiriske strategi	5
5 Empiriske resultater	8
6 Oppsummering og konklusjon	12
7 Begrensninger og mulige utvidelser	12
8 Kilder	12
9 Appendix	13

1 Innledning

I denne avhandlingen skal jeg analysere betydningen av å gå i barnehage for leseferdighetene til barn i 4 klasse.

Det er flere forhold som gjør dette spørsmålet interessant å se på. For det første er det veldig mange yrkesaktive og behovet for og bruken av barnehage har økt voldsomt i de senere årene. I tillegg er det økt fokus på utdanning og god kompetanse for bedre verdiskapning i samfunnet. Vi får og et spørsmål om vi må endre på barnehagesystemet vi har i dag for å forbedre det eller om det fungerer.

I utgangspunktet har jeg valgt å starte med hypotesen at det å gå i barnehage har en positiv påvirkning på leseresultatene, fordi jeg har en antagelse om at barnehagene driver med en del aktiv læring og en lærer seg sosiale normer som kan være nyttige. Jeg vil i oppgaven se på tidligere forskning på temaet (2), redegjøre for datagrunnlag i denne undersøkelsen (3) og beskrive den regresjonsmodellen som er laget ved hjelp av datagrunnlaget for å gjøre min egen undersøkelse (4). Deretter vil jeg teste hypotesen om at det er en forskjell på leseferdighetene i fjerde klasse mellom de som har gått i barnehage og de som ikke har (5). Ut fra dette konkluderer jeg i denne oppgaven med at det er en forskjell, og at effekten av å gå i barnehage er positiv (6).

2 Tidligere studier

Tidligere studier som ser på effekten av barnehage i relasjon til kompetanse og utdanning bruker mange ulike metoder og har forskjellige mål. De knytter seg til ulike områder og land, men de fleste undersøkelsene er fra USA. Jeg vil starte

med å se på en amerikansk studie, for så å se på to europeiske.

I den amerikanske studien sammenligner de leseevnen til barn i barnehage med leseevnen til hjemmelærte barn, og bruker data fra 60 deltagere. Denne studien konkluderer med at "In contrast to previous research, our results indicated significantly lower early literacy skills among the homeschooled children compared to the formally schooled kindergartners". En svakhet med denne studien er imidlertid at antall observasjoner er liten, som gir noe usikkerhet i resultatet. (Aram, D., et al., (2016))

I Europa er det desverre få studier som spesifikt sammenligner det å ha gått i barnehage med det å ikke ha gått i barnehage. Dette kan være grunnet at systemene er mer homogene i de enkelte europeiske landene og at det derfor har vært mindre behov eller interesse for å gjennomføre studier. Jeg har valgt å ta med en norsk og en tysk studie, som er noen av de relativt få jeg har funnet for Europa, som ikke går på leseevne, men som allikevel kan belyse positive læringseffekter av det å gå i barnehage.

Den tyske studien tar for seg barnehager vest i Tyskland, og har et utvalg på cirka 5000 deltakere. Den behandler to hovedspørsmål, den ser i utgangspunktet på sammenhengen mellom det å gå i barnehage og innplassering i 7 klasse i det Tyske skolesystemet, men differensierer også mellom barn av Tyske statsborgere og barn med innvandrerbakgrunn. Undersøkelsen konkluderer med at det for barn med innvandrerbakgrunn er en sannsynlighet for at man flyttes til et høyere utdanningsnivå når man har gått i barnehage. Det påvises ikke noen signifikant forskjell for barn av Tyske statsborgere, men det utelukkes ikke at det kan være en forskjell. En svakhet det pekes på i den Tyske undersøkelsen er at den bare baserer seg på data fra vest i Tyskland, men det er ikke helt åpenbart at det er en svakhet som er viktig som bakgrunn for min undersøkelse. (Spiess, C., et al., (2003))

Den siste undersøkelsen er en norsk studie, som ser på langtidseffekter av det å gå i barnehage og som konkluderer med at "Barnehage har store positive effekter på barnas utdanningsnivå og arbeidsmarkedstilknypning i voksen alder". I tillegg pekes det på at det er barn med innvandrerbakgrunn som har best utbytte av barnehagen. (Havnes, T., Mogstad, M., (2009))

De undersøkelsen som er gjennomgått ovenfor gir ikke grunnlag for sikre konklusjoner, fordi det er et begrenset utvalg og det er en del usikkerhet fordi undersøkelsene måler ulike effekter av det å gå i barnehage. Ser en de tre undersøkelsene samlet trekker det imidlertid i retning av at barnehage har en positiv effekt i en del år etter at barnehagen er ferdig. I den norske studien synes det klart at barnehagen har hatt en tydelig positiv langtidseffekt. Selv går jeg inn i oppgaven med en hypotese om at barnehage har en positiv effekt på leseferdighetene i 4 klasse.

3 Datamaterialet

For å belyse problemstillingen har jeg i hovedsak innhentet data fra en stor undersøkelse utført av "International association for the evaluation of educational achievement, IEA" , som heter Reading literacy study. Denne tar for seg leseferdighetene til elever i fjerde klasse i flere land. Undersøkelsen som ble gjort i 2001 tar for seg 35 land og omfatter totalt ca. 150000 elever. I tillegg til testresultatene ble det spurt om flere andre bakgrunns spørsmål for å kartlegge klassestørrelse, foreldrenes bakgrunn og lærernes bakgrunn. Under (i table 1) er en oversikt over alt det ble spurt om og de tilhørende variablene. Disse variablene kan deles i tre undergrupper, det vil si dummyvariabler kategorivariabler og kontinuerlige variabler. Dummyvariabler er variabler hvor en bare har to kategorier for eksempel hankjønn og hunkjønn. Kategorivariablene er variabler hvor en har mer enn to kategorier som for eksempel ulike inntektsnivåer. De kontinuerlige variablene har verdien de har i virkeligheten og er ofte variabler som år med erfaring, og år med utdanning. En oversikt over inndelingen av de forskjellige variablene ligger i datasettet for undersøkelsen.

<i>read</i>	Testscore	kontinuerlig
<i>idgrade</i>	Hvilken klasse resultatet er fra	kategorivariabel
<i>birthy</i>	året eleven er født	kategorivariabel
<i>brithm</i>	måneden eleven er født	kategorivariabel
<i>girl</i>	kjønn	dummy
<i>not_born</i>	Var eleven født i landet eller ikke	dummy
<i>speak_testlanguage</i>	Hvor mye snakker eleven språket i landet hjemme	kategorivariabel
<i>books_home</i>	antall bøker hjemme	kategorivariabel
<i>early_ability</i>	hvor god var eleven tidlig i språk	kategorivariabel
<i>kinderg_att</i>	Gikk eleven i barnehage	dummy
<i>par_not_born</i>	Er foreldrene født i landet	dummy
<i>par_edu</i>	foreldrenes høyeste fulførte utdanning	kategorivariabel
<i>par_emp</i>	Foreldrenes jobbstatus	kategorivariabel
<i>income</i>	Gjennomsnittlig inntekt i husstanden	kategorivariabel
<i>cl_size</i>	Anntall studenter i klassen	kontinuerlig
<i>teacher_exp</i>	Antall år læreren har jobbet	kontinuerlig
<i>teacher_fem</i>	Er læreren kvinnelig	dummy
<i>teacher_age</i>	Lærerens alder	kontinuerlig
<i>teacher_cert</i>	Har læreren sertifikasjon	dummy
<i>teacher_edu</i>	Har læreren utdanning på universitetsnivå	dummy
<i>pc_class</i>	Har klassen tilgang på pc	dummy
<i>schoolsize4</i>	Antall fjerdeklasseelever på skolen	kontinuerlig
<i>school_location</i>	Antall inbyggere hvor skolen er lokalisert	kategorivariabel
<i>pct_disadv</i>	Prosent elever som kommer fra økonomisk utsatte familier	kategorivariabel
<i>pct_abroad</i>	prosent elever født i et annet land på skolen	kategorivariabel
<i>sameteacher_4plus</i>	elever på skoler hvor de normalt har samme lærer i 4 år	dummy
<i>sameteacher_1less</i>	elever på skoler hvor de normalt har samme lærer i 1 år	dummy

Table 1: Oversikt over variabler

Jeg har i oppgaven valgt å se på data fra England, Norge og Argentina. Grunnen til det er at datamaterialet virket bra, og at en får analysert data fra flere kulturer. For å få en bedre oversikt over materialet ser jeg på fem parametre, det vil si gjennomsnitt, standardavvik, min, max og antall observasjoner. Gjennomsnittet er gjennomsnittet av verdiene til variabelen, standardavviket er forventet avvik fra gjennomsnittet (kvadratroten av variansen), max og min er høyeste og laveste verdi funnet i dataen og obs er antall observasjoner av dataen. Siden jeg i oppgaven har bestemt meg for å bruke data fra England, Norge og Argentina er det data fra disse landene som er listet i tabellen. (Table 2)

	Gjennomsnitt	Standardavvik	Min	Max	Obs
<i>read</i>	490.443	98.5753	161.4514	773.3768	9,915
<i>clsize</i>	25.81873	6.300476	4	39	9,461
<i>income</i>	3.797432	1.604194	1	6	4,517
<i>teacher_exp</i>	14.98387	10.1618	1	42	9,362
<i>books_home</i>	3.25995	1.411673	1	5	7,186
<i>par_edu</i>	2.684195	1.471696	1	5	6,732
<i>par_emp</i>	1.888164	.7965441	1	4	6,134
<i>pct_abroad</i>	1.13194	.4239517	1	4	8,898
<i>pct_disadv</i>	2.271518	1.239527	1	4	9,097
<i>schoolsize4</i>	62.27866	51.42517	3	586	9,237
<i>speak_testlang_home</i>	1.17631	.4541691	1	3	9,177
<i>early_ability</i>	2.71018	.9312032	1	4	7,063
<i>not_born</i>	.1948481	.3961055	0	1	9,084
<i>sameteahcer1_less</i>	.4369093	.4960294	0	1	9,629
<i>sameteacher4_plus</i>	.144771	.3518881	0	1	9,629
<i>school_location</i>	2.379949	.8652066	1	4	7,830

Table 2: Oversikt over verdier på interessevariabler

4 Empiriske strategi

I denne seksjonen skal jeg ved hjelp av regresjonsanalyse finne en kurve som modellerer dataen fra datasettet. Regresjonsanalyse er en analysemetode som beskriver sammenhengen mellom én eller flere uavhengige variabler og en avhengig variabel.

Når jeg utfører regresjonsanalysen er det viktig å ta i betraktning alle forhold som kan påvirke resultatet slik at modellen blir best mulig. Jeg vil gå igjennom de enkelte variablene jeg vil ta hensyn til og begrunne generelt hvorfor det kan være viktig å ta hensyn til disse. Til slutt vil jeg konstruere testen som jeg skal bruke for å se på forskjellene i leseferdigheter mellom å ha gått i barnehage og det å ikke ha gjort det.

For å få et godt inntrykk av hvilke variabler jeg skal ta hensyn til kan det være fornuftig å lage en korrelasjonsmatrise for å få en oversikt over sammenhengen mellom variablene (se Appendix). Jeg er da spesielt interessert i korrelasjonen mellom *read* variabelen (leseferdigheter) og de andre variablene. Ved å bruke matrisen og forventninger kan jeg lage en modell som treffer dataen ganske godt.

For å finne frem til en best mulig modell lager jeg tre ulike modeller for sammenligning og vurdering av hvilken modell som tilsynelatende gir mest pålitelig resultat. Jeg starter med en grunnmodell (1). I modell 1 har jeg tatt med noen av de parameterne jeg antar er viktigst. Jeg forventer at klassestørrelse er viktig, siden mer tid per elev bør ha en positiv innvirkning på leseresultatet. En bedre inntekt i familien kan ofte bety bedre skole eller privat leksehjelp og konsekvensen av dette er et bedre leseresultat. Tilgang på bøker, lærerens

erfaring og tidlige evner hos barna er andre ting som jeg forventer gir bedre leseferdigheter. Modell 1 får da følgende elementer:

$$(1)read = \beta_0 + \beta_1clsize + \beta_2income + \beta_3teacher_exp + \beta_4books_home + \beta_5early_ability + \epsilon_i$$

Modell (1) utelater flere andre aspekter som kan være viktige, og en del av disse inkluderes i modell 2 og 3.

I modell 2 beholdes elementene fra modell 1 og så legger jeg til en del andre viktige elementer blant annet språk en snakker hjemmet, om en er født i landet, foreldrenes utdanning og hvor mange prosent i klassen som kommer fra økonomisk utsatte familier. Modell 2 blir da som følger:

$$(2)read = \beta_0 + \beta_1clsize + \beta_2income + \beta_3teacher_exp + \beta_4books_home + \beta_5early_ability + \beta_6par_emp + \beta_7pct_disadv + \beta_8schoolsize4 + \beta_9speak_testlang_home + \beta_{10}par_edu + \beta_{11}not_born + \epsilon_i$$

I modell 3 vil jeg se hva som skjer hvis jeg bruker ikke linearitet i størrelse på klassen. Grunnen til det er at tid per elev endrer seg mye mer når en går fra fire til fem sammenlignet med satten til atten, og kan derfor trolig beskrives bedre av en ikkelineær modell. Med utgangspunkt i modell 2 legger jeg til variabelen $clsize_sq$ som er gitt ved $clsize_sq = clsize^2$

$$(3)read = \beta_0 + \beta_1clsize + \beta_2income + \beta_3teacher_exp + \beta_4books_home + \beta_5early_ability + \beta_6par_emp + \beta_7pct_disadv + \beta_8schoolsize4 + \beta_9speak_testlanguage + \beta_{10}par_edu + \beta_{11}not_born + \beta_{12}clsize_sq + \epsilon_i$$

Resultatet fra de tre forskjellige regresjonsmodellene er gitt under i table3. Her er verdien for estimatet av variablene $\beta_0 \dots \beta_{12}$ for modell 1,2 og 3. Standardavviket er gitt i parantes.

	(1)	(2)	(3)
Variables	read	read	read
<i>clsize</i>	3.803596 (.1786862)	3.651154 (.2238141)	-7.495185 (1.227042)
<i>income</i>	5.359204 (.7545461)	3.246944 (.9713055)	3.562202 (.9595842)
<i>teacher_exp</i>	-.2322415 (.1019848)	-.4111221 (.1132007)	-.4105681 (.1117639)
<i>books_home</i>	12.5401 (1.074089)	8.297258 (1.283569)	8.728975 (1.268139)
<i>early_ability</i>	32.19227 (1.199467)	31.27553 (1.328608)	30.70261 (1.31321)
<i>par_emp</i>		.1887394 (1.9772)	.0629882 (1.952151)
<i>pct_disadv</i>		8.627623 (1.870153)	8.255777 (1.846855)
<i>schoolsize4</i>		.016699 (.0540032)	.1039573 (.0541485)
<i>speak_testlang_home</i>		-24.9531 (3.796489)	-24.07 (3.74952)
<i>par_edu</i>		-7.919722 (1.179127)	-8.477918 (1.165729)
<i>not_born</i>		-29.63488 (4.477951)	-29.38911 (4.421193)
<i>clsize_sq</i>			.23641(.0255996)
<i>constant</i>	281.5341 (7.069517)	353.1472 (11.41465)	470.029 (16.94684)
<i>Observations</i>	4,267	3,269	3,269
<i>R – sqared</i>	0.2702	0.3092	0.3268

Table 3: Tre ulike regresjonsmodeller

Jeg ser ut fra disse at de fleste resultatene er slik jeg forventet, at leseferdighetene øker med økt inntekt, flere bøker i hjemmet og gode evner tidlig. Det mest overraskende er at lærerens erfaring i alle tre modeller har en litt negativ effekt, men en mulig forklaring på det kan være at motivasjonen kan være høyest når en er ferdig med utdanningen som lærer og det er da en har de pedagogiske fagene best i minne. Ser en på *par_emp* (foreldrenes jobbstatus) ser en at antatt verdi er nesten null. Ved bruk av en T-test kan jeg teste sannsynligheten for at denne verdien er lik null, det vil si at den ikke har noen innvirkning, og resultatet er 99,7%. Dette er så stor sannsynlighet for at *par_emp* ikke har noen innvirkning at jeg med stor sikkerhet kan si at den ikke påvirker leseresultatet. Nederst på resultatlisten har vi og tallet *R – sqared*, som sier noe om prosenten av observasjonene modellen jeg har generert treffer på. *R – sqared* vil alltid være mellom 0 og 1, og dess høyere verdien er dess bedre er estimatet.

Jeg skal nå lage testen som skal sammenligne de som har gått i barnehage og de som ikke har gått i barnehage. Jeg innfører derfor versjoner av de tidligere variablene som bare inneholder data fra de som har gått i barnehage. Disse variablene er definert som $variabel_kinderg = variabel \cdot kinderg_att$. Deretter lager jeg interaksjonsvariabler mellom de to, vi får derfor:

$$(4.1) read = \beta_0 + \beta_1 clsize + \beta_2 income + \beta_3 teacher_exp + \beta_4 books_home + \beta_5 early_ability + \beta_6 pct_disadv + \beta_7 schoolsize4 + \beta_8 speak_testlang_home + \beta_9 par_edu + \beta_{10} not_born + \beta_{11} clsize_sq + \delta_0 kinderg_att + \delta_1 clsize_kinderg + \delta_2 income_kinderg + \delta_3 teacher_exp_kinderg + \delta_4 books_home_kinderg + \delta_5 early_ability_kinderg + \delta_6 pct_disadv_kinderg + \delta_7 schoolsize4_kinderg + \delta_8 speak_testlang_home_kinderg + \delta_9 par_edu_kinderg +$$

$$\delta_{10}not_born_kinderg + \delta_{11}clsiz_sq_kinderg + \epsilon_i$$

Vi får da at likningen for de i barnehage er:

$$(4.2) \textit{read} = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)clsiz + (\beta_2 + \delta_2)income + (\beta_3 + \delta_3)teacher_exp + (\beta_4 + \delta_4)books_home + (\beta_5 + \delta_5)early_ability + (\beta_6 + \delta_6)pct_disadv + (\beta_7 + \delta_7)schoolsize4 + (\beta_8 + \delta_8)speak_testlang_home + (\beta_9 + \delta_9)par_edu + (\beta_{10} + \delta_{10})not_born + (\beta_{11} + \delta_{11})clsiz_sq + \epsilon_i$$

Formelen for de som ikke går i barnehage blir da:

$$(4.3) \textit{read} = \beta_0 + \beta_1clsiz + \beta_2income + \beta_3teacher_exp + \beta_4books_home + \beta_5early_ability + \beta_6par_emp + \beta_7pct_disadv + \beta_8schoolsize4 + \beta_9speak_testlang_home + \beta_{10}par_edu + \beta_{11}not_born + \epsilon_i$$

Jeg setter så opp hypotesen min,

$$(5) H_0 : \delta_1 = \delta_2 = \dots = \delta_{11} = 0 \quad H_1 : \delta_1 \neq 0 \textit{ eller } \delta_2 \neq 0 \textit{ eller } \dots \textit{ eller } \delta_{12} \neq 0$$

5 Empiriske resultater

Jeg kan nå teste hypotesen min og starter med å gjennomføre en regresjonsanalyse basert på formelen (4.1) jeg har satt opp tidligere. Resultatene fremkommer i table 4.

Wage	coef.	Std. Err.	t	$P > t $	[95 Conf. Interval]
<i>clsize</i>	-13.74254	3.521624	-3.90	0.000	-20.64738 -6.837704
<i>income</i>	1.232674	2.84311	0.43	0.665	-4.341804 6.807152
<i>teacher_exp</i>	-0.3286084	0.3168982	-1.04	0.300	-0.9499497 0.2927328
<i>books_home</i>	16.9932	3.313261	5.13	0.000	10.4969 23.4895
<i>early_ability</i>	26.92974	3.661427	7.35	0.000	19.75079 34.10868
<i>par_emp</i>	3.385973	5.609865	0.60	0.546	-7.613271 14.38522
<i>pct_disadv</i>	-0.6060683	5.53225	-0.11	0.913	-11.45313 10.241
<i>schoolsize4</i>	-0.0551622	0.1620312	-0.34	0.734	-.3728563 .2625318
<i>speak_testlang_home</i>	-30.21409	8.939474	-3.38	0.001	-47.74169 -12.68649
<i>par_edu</i>	-10.52428	3.283468	-3.21	0.001	-16.96217 -4.086399
<i>not_born</i>	-31.19288	12.70375	-2.46	0.014	-56.10107 -6.28468
<i>clsize_sq</i>	0.3933699	0.0734329	5.36	0.000	.2493902 .5373496
<i>kinderg_att</i>	-66.7168	51.78844	-1.29	0.198	-168.2582 34.82463
<i>clsize_kinderg</i>	7.056835	3.757921	1.88	0.060	-.3113098 14.42498
<i>income_kinderg</i>	2.810023	3.021563	0.93	0.352	-3.114346 8.734393
<i>teacher_exp_kinderg</i>	-0.0772379	0.338785	-0.23	0.820	-.7414927 .5870169
<i>books_home_kinderg</i>	-9.598503	3.588313	-2.67	0.008	-16.6341 -2.56291
<i>early_ability_kinderg</i>	4.019239	3.923855	1.02	0.306	-3.674252 11.71273
<i>par_emp_kinderg</i>	-3.422113	5.984813	-0.57	0.567	-15.15652 8.312292
<i>pct_disadv_kinderg</i>	9.552495	5.870813	1.63	0.104	-1.95839 21.06338
<i>schoolsize4_kinderg</i>	0.1875787	0.1719568	1.09	0.275	-.1495765 .5247339
<i>speak_testlang_home_kinderg</i>	9.516556	9.857663	0.97	0.334	-9.811332 28.84444
<i>par_edu_kinderg</i>	2.429758	3.514011	0.69	0.489	-4.460153 9.319669
<i>not_born_kinderg</i>	1.54672	13.55582	0.11	0.909	-25.03214 28.12558
<i>clsize_sq_kinderg</i>	-0.1781958	0.0783696	-2.27	0.023	-.3318549 -.0245367
constant	527.1155	48.51183	10.87	0.000	431.9986 622.2325

Table 4: Regresjonsanalyse av likning (4.1)

Jeg kan nå gjøre en Chow-test på resultatet for å teste om $\delta_1 = \delta_2 = \dots = \delta_{11} = 0$, altså teste om min hypotese (5) stemmer. For å gjøre dette trenger jeg både modellen for likning (3) og (4.1). Jeg lager derfor en samlet tabell for begge modellene. (Se figur 5)

	(3)	(4.1)
Variables	read	read
<i>clsize</i>	-7.495185 (1.227042)	-13.74254 (3.521624)
<i>income</i>	3.562202 (0.9595842)	1.232674 (2.84311)
<i>teacher_exp</i>	-0.4105681 (0.1117639)	-0.3286084 (0.3168982)
<i>books_home</i>	8.728975 (1.268139)	16.9932 (3.313261)
<i>early_ability</i>	30.70261 (1.31321)	26.92974 (3.661427)
<i>par_emp</i>	0.0629882 (1.952151)	3.385973 (5.609865)
<i>pct_disadv</i>	8.255777 (1.846855)	-0.6060683 (5.53225)
<i>schoolsize4</i>	0.1039573 (.0541485)	-0.0551622 (0.1620312)
<i>speak_testlang_home</i>	-24.07 (3.74952)	-30.21409 (8.939474)
<i>par_edu</i>	-8.477918 (1.165729)	-10.52428 (3.283468)
<i>not_born</i>	-29.38911 (4.421193)	-31.19288 (12.70375)
<i>clsize_sq</i>	0.23641(0.0255996)	0.3933699 (0.0734329)
<i>kinderg_att</i>		-66.7168 (51.78844)
<i>clsize_kinderg</i>		7.056835 (3.757921)
<i>income_kinderg</i>		2.810023 (3.021563)
<i>teacher_exp_kinderg</i>		-0.0772379 (0.338785)
<i>books_home_kinderg</i>		-9.598503 (3.588313)
<i>early_ability_kinderg</i>		4.019239 (3.923855)
<i>par_emp_kinderg</i>		-3.422113 (5.984813)
<i>pct_disadv_kinderg</i>		9.552495 (5.870813)
<i>schoolsize4_kinderg</i>		0.1875787 (0.1719568)
<i>speak_testlang_home_kinderg</i>		9.516556 (9.857663)
<i>par_edu_kinderg</i>		2.42975 (3.514011)
<i>not_born_kinderg</i>		1.54672 (13.55582)
<i>clsize_sq_kinderg</i>		-0.1781958 (0.0783696)
constant	470.029 (16.94684)	527.1155 (48.51183)
Observations	3,269	3,264
R-squared	0.3268	0.3322

Table 5: Samling av modell (3) og (4.1)

Jeg kan nå sette opp formelen for Chow-testen som er gitt ved:

$$TS = \frac{R_w^2 - R_d^2}{\frac{h}{1 - R_w^2}} \sim F_{h, n-k}$$

R_w^2 =R-squared verdi for likning (4.1) (se tabell 5)

R_d^2 =R-squared verdi for likning (3) (se tabell 5)

(n-k)=antall frihetsgrader i den urestriktete modellen

h=antall restriksjoner

Antall frihetsgrader er gitt ved hvor mange variabler en fritt kan variere i modellen. Jeg velger å bruke en p-verdi på 0.05 som er normalt for de fleste undersøkelser. Jeg regner så ut verdien på Chow-testen for så å sammenligne med den kritiske verdien.

$$TS = \frac{\frac{0.3322-0.3268}{13}}{\frac{1-0.3322}{3,264-25-1}} = 2.01$$

Den kritiske verdien er hentet fra tabell for Chow-test og er

$$F(13, 3238, 0.05)^{-1} = 1.7231756 < TS$$

Jeg finner en signifikant forskjell på leseferdighetene i fjerde klasse mellom å gå i barnehage og ikke å gjøre det.

Til slutt skal jeg prøve å finne ut om endringen er positiv eller negativ, det vil si om barnehage har positiv eller negativ effekt på leseferdighetene i fjerde klasse. Jeg vil derfor sammenligne modellen for de i barnehage med modellen for alle elever generelt. Dette gjør en ved å sette inn gjennomsnittsverdier fra tabellen 1 i modellene (3) og (4.2) for så å regne ut verdien på lesetesten.

Jeg starter med å løse likning 3 ved hjelp av tabell 1.

$$(3) read = \beta_0 + \beta_1 clsize + \beta_2 income + \beta_3 teacher_exp + \beta_4 books_home + \beta_5 par_edu + \beta_6 par_emp + \beta_7 pct_disadv + \beta_8 schoolsize4 + \beta_9 speak_testlanguage + \beta_{10} early_ability + \beta_{11} not_born + \beta_{12} clsize_sq + \epsilon_i = \beta_0 + \beta_1 \cdot 25.82 + \beta_2 \cdot 3.80 + \beta_3 \cdot 14.98 + \beta_4 \cdot 3.26 + \beta_5 \cdot 2.68 + \beta_6 \cdot 1.89 + \beta_7 \cdot 1.13 + \beta_8 \cdot 62.28 + \beta_9 \cdot 1.18 + \beta_{10} \cdot 2.71 + \beta_{11} \cdot 0.19 + \beta_{12} \cdot (25.82)^2 = 511.2$$

Jeg fyller så inn variablene $\delta_1 = \delta_2 = \dots = \delta_{11}$ og $\beta_1 = \beta_2 = \dots = \beta_{11}$, fra tabell 4 for å løse likning (4.2).

$$(4.2) read = (\beta_0 + \delta_0) + (\beta_1 + \delta_1) clsize + (\beta_2 + \delta_2) income + (\beta_3 + \delta_3) teacher_exp + (\beta_4 + \delta_4) books_home + (\beta_5 + \delta_5) early_ability + (\beta_6 + \delta_6) pct_disadv + (\beta_7 + \delta_7) schoolsize4 + (\beta_8 + \delta_8) speak_testlang_home + (\beta_9 + \delta_9) par_edu + (\beta_{10} + \delta_{10}) not_born + (\beta_{11} + \delta_{11}) clsize_sq + \epsilon_i = (\beta_0 + \delta_0) + (\beta_1 + \delta_1) \cdot 25.82 + (\beta_2 + \delta_2) \cdot 3.80 + (\beta_3 + \delta_3) \cdot 14.98 + (\beta_4 + \delta_4) \cdot 3.26 + (\beta_5 + \delta_5) \cdot 2.68 + (\beta_6 + \delta_6) \cdot 1.89 + (\beta_7 + \delta_7) \cdot 1.13 + (\beta_8 + \delta_8) \cdot 62.28 + (\beta_9 + \delta_9) \cdot 1.18 + (\beta_{10} + \delta_{10}) \cdot 2.71 + (\beta_{11} + \delta_{11}) \cdot 0.19 + (\beta_{12} + \delta_{12}) \cdot (25.82)^2 = 513.85$$

Ut fra ovenstående beregning ser vi at en får et høyere gjennomsnittresultat for de som har gått i barnehage sammenlignet med de som ikke har, hvilket betyr at det å gå i barnehage har en positiv effekt på leseferdighetene. Det er imidlertid viktig å poengtere at det å bruke gjennomsnittsverdier ikke sammenligner hele modellen og dette er en svakhet som må tas i betraktning.

6 Oppsummering og konklusjon

I denne oppgaven startet jeg med å se på tidligere forskning, og tok utgangspunkt i en hypotese om at det å gå i barnehage har en positiv effekt på leseferdighetene i fjerde klasse. I gjennomgangen av tidligere forskning har jeg vurdert det slik at denne trekker i retning av at barnehage har en positiv effekt i en del av ettersom at barnehagen er ferdig. Deretter har jeg sett nærmere på dataene brukt i oppgaven, som er en stor internasjonal studie med 150000 elever over 35 land, hvor jeg har valgt å se på Norge, England og Argentina. Jeg gikk så over til å lage en modell for dataene som er samlet inn og definere hypotesen matematisk. Så gjennomførte jeg testen, hvor jeg fant en statistisk signifikant forskjell på leseferdighetene i fjerde klasse mellom å gå i barnehage og ikke å gjøre det. Til slutt fant jeg at endringen var positiv.

Hovedkonklusjonen er derfor at undersøkelsen bekrefter hypotesen om at det å gå i barnehage har en positiv effekt på leseferdighetene i fjerde klasse når en sammenligner med det å ikke gå i barnehage.

7 Begrensninger og mulige utvidelser

Det er to begrensninger i oppgaven som jeg vil poengtere. For det første er dataen snart 20 år gamle og reflekterer ikke nødvendigvis dagens system. Det er også slik at når systemene varierer over landegrensener er det vanskelig å gjøre generell forskning på viktigheten av barnehager fordi påvirkningen kan variere fra land til land.

Det er flere utvidelser som er mulig basert på disse resultatene. En av de er å bruke nyere data for å få et bedre bilde av det nåværende systemet. Det hadde også vært interessant å gjøre en undersøkelse som følger deltakerne over lengre tid for så å se om effekten er langvarig eller er en mer kortvarig positiv effekt. I tillegg kunne en forsøke å kvantifisere effekten av det å gå i barnehage på leseevnen.

8 Kilder

Aram, D., Meidan, I. C., Deitcher, D. B., (2016). A comparison between homeschooled and formally schooled kindergartners: Children's early literacy, mothers' beliefs, and writing mediation. *Reading Psychology*, 37(7), 995-1024

Spiess, C. K., Büchel, F., Wagner, G. G. (2003). Children's school placement in Germany: Does Kindergarten attendance matter? *Early Childhood Research Quarterly*, 18, 255-270.

Havnes, T., Mogstad, M., (2009). Er barnehage bra for barna?, økonomiske analyser, 42-49

9 Appendix

Corrolation	<i>read</i>	<i>not_born</i>	<i>par_not_born</i>	<i>early_ability</i>	<i>par_edu</i>	<i>income</i>	<i>kinderg_att</i>
<i>read</i>	1						
<i>not_born</i>	-0.1420	1					
<i>par_not_born</i>	-0.0514	0.2168	1				
<i>early_ability</i>	0.3907	-0.0345	0.0305	1			
<i>par_edu</i>	-0.1272	0.0198	0.0514	-0.0173	1		
<i>income</i>	0.1127	-0.0826	-0.1448	0.0090	-0.4578	1	
<i>kinderg_att</i>	0.1102	-0.0183	-0.0318	0.0675	-0.0753	0.0848	1
<i>speak_test_home</i>	-0.1368	0.2237	0.3997	0.002	0.0704	-0.1184	-0.0432
<i>books_home</i>	0.1460	-0.0690	-0.1886	0.0287	-0.4744	0.4030	0.0278
<i>par_emp</i>	-0.0342	0.0345	0.0538	0.0082	0.2238	-0.3882	-0.0884
<i>clsiz</i>	0.2907	0.0174	0.0323	0.1154	0.1886	-0.1115	0.0202
<i>teacher_exp</i>	-0.0969	-0.0091	-0.0647	-0.0109	-0.0715	0.0641	-0.0714
<i>teacher_age</i>	-0.0871	-0.0129	-0.0882	-0.0365	-0.1004	0.0850	-0.0426
<i>teacher_fem</i>	-0.0635	0.0002	0.0125	-0.0162	-0.0497	0.0749	0.0015
<i>teacher_cert</i>	-0.0251	0.0065	-0.0486	-0.0027	-0.0042	-0.0427	-0.0044
<i>pc_class</i>	0.0528	0.0038	0.0170	0.0304	0.1113	-0.0706	0.0391
<i>schoolsize4</i>	0.1526	-0.0059	0.0799	0.0368	0.0598	-0.0055	0.0205
<i>school_location</i>	0.1355	0.0231	0.1508	0.0441	-0.0018	0.0150	0.0519
<i>pct_disadv</i>	0.1236	0.0765	0.1275	0.1105	0.3024	-0.3227	0.0566
<i>pct_abroad</i>	0.0279	0.0898	0.2651	0.0136	0.0275	-0.0908	-0.0130
<i>sameteacher4_plus</i>	-0.1734	-0.0525	-0.0662	-0.0633	-0.1404	0.1382	0.0118
<i>sameteacher1_less</i>	0.3976	0.0299	0.0767	0.1682	0.3261	-0.2283	0.0635

Corrolation	<i>speak_test_home</i>	<i>books_home</i>	<i>par_emp</i>	<i>clsiz</i>	<i>teacher_exp</i>	<i>teacher_age</i>
<i>speak_test_home</i>	1					
<i>books_home</i>	-0.1157	1				
<i>par_emp</i>	0.0421	-0.1667	1			
<i>clsiz</i>	0.0311	-0.1497	0.0804	1		
<i>teacher_exp</i>	-0.0257	0.0207	0.0042	-0.1272	1	
<i>teacher_age</i>	-0.0288	0.0688	-0.0252	-0.1695	0.8167	1
<i>teacher_fem</i>	0.0118	0.0415	-0.0380	-0.1559	0.0099	0.1310
<i>teacher_cert</i>	-0.0071	0.0165	-0.0012	-0.0776	0.0844	0.1101
<i>pc_class</i>	-0.0191	-0.1026	0.0271	0.1843	-0.2132	-0.2297
<i>schoolsize4</i>	0.0662	-0.0546	0.0258	0.4617	-0.0559	-0.0774
<i>school_location</i>	0.0678	-0.0285	-0.0269	0.2502	-0.1617	-0.2482
<i>pct_disadv</i>	0.0915	-0.2267	0.1393	0.3096	-0.2022	-0.2499
<i>pct_abroad</i>	0.1304	-0.0546	0.0266	0.1229	-0.1239	-0.1075
<i>sameteach4_plus</i>	-0.0509	0.1129	-0.1023	-0.3208	-0.0293	0.0430
<i>sameteach1_less</i>	0.0495	-0.2486	0.1658	0.6210	-0.1407	-0.2424

Corrolation	<i>teacher_fem</i>	<i>teacher_cert</i>	<i>pc_class</i>	<i>schoolsize4</i>	<i>school_location</i>
<i>teacher_fem</i>	1				
<i>teacher_cert</i>	-0.0723	1			
<i>pc_class</i>	-0.1109	-0.0627	1		
<i>schoolsize4</i>	-0.2882	-0.1119	0.1251	1	
<i>school_location</i>	-0.1012	-0.0764	0.2091	0.3589	1
<i>pct_disadv</i>	-0.0970	0.0660	0.1538	0.0885	0.2200
<i>pct_abroad</i>	0.0181	0.0060	-0.0151	0.0628	0.3409
<i>sameteach4_plus</i>	0.1512	0.0050	0.0389	-0.1391	-0.0228
<i>sameteach1_less</i>	-0.1568	-0.0387	0.2281	0.3254	0.2400

Corrolation	<i>pct_disadv</i>	<i>pct_abroad</i>	<i>sameteach4_plus</i>	<i>sameteach1_less</i>
<i>pct_disadv</i>	1			
<i>pct_abroad</i>	0.2200	1		
<i>sameteach4_plus</i>	-0.2260	-0.1917	1	
<i>sameteach1_less</i>	0.5003	0.0689	-0.4152	1