

Brita Engen Finne

SØK2901 Bacheloroppgave i samfunnsøkonomi

Analyse av betydningen av å ha gått i barnehage
på elevprestasjoner

Bacheloroppgave i MSØK/5

Veileder: Bjarne Strøm

Mai 2020

SØK2901 Bacheloroppgave i samfunnsøkonomi

Analyse av betydningen av å ha gått i barnehage på elevprestasjoner

14. mai 2020

Sammendrag

Formålet med oppgaven er å undersøke om det å gå i barnehage har betydning for leseprestasjonene til fjerdeklassinger i Norge. I tillegg tar den for seg hvorvidt effekten av å gå i barnehage avhenger av foreldrenes bakgrunn med tanke på innvandrerstatus og grad av utdanning. Det kontrolleres også for andre faktorer som tenkes å kunne påvirke testresultatet.

For å gjennomføre analysen er estimeringsmetoden OLS, minste kvadrats metode, benyttet og modellen som brukes er tatt med utgangspunkt i skoleproduktfunksjonen.

Resultatet viser en positiv sammenheng mellom barnehagedeltakelse og økte leseresultater blant fjerdeklassinger i Norge. Samtidig vises det gjennom kontrollvariablene at det også er en rekke andre faktorer som har stor betydning for prestasjonen. Den positive effekten av å gå i barnehage viser seg også å være større blant barn med foreldre som ikke er født i Norge.

Innholdsfortegnelse

Sammendrag	i
Innholdsfortegnelse	ii
Tabelloversikt	iii
Figuroversikt	iii
1 Introduksjon	1
1.1 Motivasjon.....	1
1.2 Problemstilling	1
2 Teoretisk rammeverk og tidligere litteratur	1
2.1 Innledning.....	1
2.2 Teoretisk rammeverk.....	1
2.3 Tidligere litteratur og resultater.....	2
2.4 Oppsummering	3
3 Datamateriale	3
3.1 Innledning.....	3
3.2 Definisjoner og kilder.....	3
3.3 Deskriptiv statistikk for avhengig variabel	4
3.4 Deskriptiv statistikk for interessevariabel	4
3.5 Deskriptiv statistikk for kontrollvariablene.....	6
3.6 Korrelasjonsmatrise for datamaterialet	8
3.7 Oppsummering	9
4 Økonometrisk modell	9
4.1 Innledning.....	9
4.2 Valg av funksjonsform	9
4.3 Empirisk strategi og estimeringsmetode	12
4.4 Oppsummering	13
5 Empiriske resultater	14
5.1 Innledning.....	14
5.2 Empiriske hovedresultater.....	14
5.3 Empiriske resultater av tilleggsspørsmål.....	15
5.4 Diskusjon.....	17
5.5 Oppsummering	18
6 Konklusjon	18
7 Videre arbeid	19
Referanser	20
Appendix	21

Tabelloversikt

3.1 Definisjon av variabler	3
3.2 Deskriptiv statistikk for avhengig variabel	4
3.3 Deskriptiv statistikk for interessevariabel	5
3.4 Deskriptiv statistikk for avhengig variabel – gått i bhg. vs. ikke bhg.....	5
3.5 Videre variabelforklaring for utdanningsvariabelen	6
3.6 Videre variabelforklaring for årlig familieinntekt.....	7
3.7 Deskriptiv statistikk for kontrollvariablene.....	7
3.8 Deskriptiv statistikk for kontrollvariablene – gått i bhg. vs. ikke bhg.	7
3.9 Korrelasjonsmatrise for datamaterialet	8
4.1 Forkortelse av variabelnavn	9
5.1 Presentasjon av OLS-regresjon med og uten kontrollvariabler.....	14
5.2 Presentasjon av modeller inkludert OLS-regresjon med interaksjonsledd.	16

Figuroversikt

3.1 Histogram som viser fordeling av leseferdigheter i Norge	4
3.2 Histogram som viser fordelingen av prestasjonsresultater i de to gruppene.....	6
A.1 Stata-resultater for likning (3.1)	21
A.2 Stata-resultater for likning (3)	21
A.3 Stata-resultater for likning (4)	22
A.4 Stata-resultater for likning (5)	22

1 Introduksjon

1.1 Motivasjon

«Gode kommunikasjonsevner og et godt utviklet språk er en viktig forutsetning for deltagelse, læring og utvikling» står det på regjeringen sine nettsider under artikkelen «Kvalitet i barnehagen» [4]. I norske barnehager har språk fått et stort fokus, og det er derfor grunn til å tro at deltakelse i barnehage kan føre til høyere elevprestasjoner på skolen. Høy barnehageoppslutning i Norge kan av denne grunn være gunstig for individet, men også samfunnsøkonomisk sett dersom dette bidrar til å øke ferdighetsnivået i befolkningen.

For å få et godt utdanningssystem og for at flest mulig skal være godt rustet til skolestart er det viktig å kartlegge betydningen av barnehagens påvirkning på barnet. Målet med oppgaven er derfor, ved hjelp av empirisk analyse, å undersøke dens betydning i forhold til andre faktorer som også kan spille inn på prestasjoner i skolen.

1.2 Problemstilling

Hvordan påvirkes leseprestasjoner av å ha gått i barnehage i Norge?

Tilleggsspørsmål:

Avhenger den eventuelle effekten av å ha gått i barnehage av foreldrenes bakgrunn med hensyn til grad av utdanning og innvandrersstatus?

2 Teoretisk rammeverk og tidligere litteratur

2.1 Innledning

Dette kapittelet gir en beskrivelse av barnehagens opprinnelse og deres betydning i samfunnet. I tillegg vil det inneholde en kort oppsummering av tidligere relevant litteratur.

2.2 Teoretisk rammeverk

Barnehagen ble grunnlagt som en pedagogisk institusjon av Friedrich Fröbel, noe som har hatt stor betydning for barnehagens utvikling i Norge og Norden. Fröbel hadde fokus på pedagogisk opplæring og arbeidet i flere år med utviklingen av en allsidig lekemodell. Fokuset på pedagogikk var i motsetning til det som i hovedsak tidligere var ment som et tiltak til de som var i direkte nød, kalt asyl. Han startet Fröbel-seminarer som utdannet pedagoger til barnehagene, hvor blant annet norske kvinner tok utdanning og startet private virksomheter i Norge [2].

I barnehagen legges grunnlaget for videre læring og utvikling. Dette er en arena hvor barna tilegner seg verdier, holdninger og praksis for et bærekraftig samfunn, og i Norge er det gått fra å være et privilegium for noen få til å være en institusjon for alle. Det ble i 2009 lovfestet

at alle som ønsker barnehageplass har rett til dette, og rundt 90 % av alle barn i Norge i dag går i barnehagen. I dag regnes den som et fundament for videre skolegang og som en del av et livslangt læringsløp [3].

Det er mange faktorer som har betydning for hvordan en elev presterer på skolen. Ofte snakker man om en klassisk «bra» eller «dårlig» skole, men hva som gjør en skole bra er vanskelig å vite. I tillegg er det mye som tyder på at elevens oppvekst og familiebakgrunn har stor betydning for videre utdanning. I denne oppgaven tas det utgangspunkt i at testresultater i skolen kan måles ut ifra en skoleproduktfunksjon [4].

$$T = f(S, F, P) \quad (0)$$

Likningen over presenterer skoleproduktfunksjonen hvor,

T = testscore

S = Skolefaktorer (Klassestørrelse, lærerkarakteristika)

F = Familie/elevkarakteristika

P = Medelevkarakteristika (peer group- effekter)

Denne modellen fanger opp en rekke tenkelige faktorer som kan ha innvirkning på utdanningsforløpet til en elev, blant annet barnehagedeltakelse, og brukes i denne oppgaven for å beregne elevens testscore.

2.3 Tidligere litteratur og resultater

Artikkelen «Education production function» [4] tar for seg faktorer som kan påvirke resultatene til elevene i en såkalt skoleproduktfunksjon og dens styrker og svakheter. En av motivasjonene ligger i at gode individuelle ferdigheter gir en fordel i arbeidsmarkedet og lønnsnivået. Viktigheten av både skoleressurser, lærerkvalitet og familieegenskaper diskuteres og forfatteren, Hanushek, påpeker at det er vanskelig å si hvilke faktorer som er mest avgjørende. Det finnes utrolig mange relevante faktorer og det er veldig omfattende og krevende å sette sammen en undersøkelse med et representativt utvalg. Likevel argumenteres det for at hvordan man bruker ressursene i skolen er viktigere enn mengden ressurser man har tilgjengelig. I tillegg ser man også at ulik kvalitet på lærer er den sterkeste signifikante forskjellen mellom skoler.

Hans Bonesrønning tar i sin artikkel, «Utforming i utdanningspolitikken» [5], opp hvorvidt økonomer kan gi råd til hvordan man kan oppnå bedre resultater og produktivitet i skolesektoren. Han påpeker at styringsformen burde ta fokuset bort fra ressursinnsatsen og heller bli mer resultatfokusert. Det vises til at man tidligere har tenkt at klassestørrelse har en stor betydning, men at det i dag er en bred enighet om at lærerens karakteristika har en viktig betydning for elevenes prestasjoner uten at det er helt klart hvilke karakteristika dette er. Forfatterens personlige mening er at det er mange små effekter som er vanskelig å fange opp i data som er den viktigste årsaken til gode resultater. Det konkluderes med at økonomer har en

viktig betydning når det kommer til hvordan dette kan organiseres spesielt på grunn av deres kunnskap om ressursallokering og insentiver til bedre resultater og effektivitet i skolen.

2.4 Oppsummering

I dette kapittelet har bakgrunnen for den norske barnehagen blitt presentert og motivasjonen for oppgaven. Det er også gjort rede for det teoretiske rammeverket for modellen det tas utgangspunkt i, skoleproduktfunksjonen, samt noe tidligere litteratur.

3 Datamaterialet

3.1 Innledning

Dette kapittelet vil ta for seg datamaterialet som oppgaven tar utgangspunkt i. Aktuell datafil som blir benyttet med forklaring av variabler blir presentert først. Deretter presenteres en deskriptiv statistikk av den avhengige variabelen, interessevariabelen og kontrollvariablene med nødvendige forklaringer. Til slutt vil datamaterialets korrelasjonsmatrise bli presentert sammen med noen forklaringer og potensielle problemer.

3.2 Definisjoner og kilder

I denne oppgaven er datasettet som benyttes hentet fra den internasjonale undersøkelsen Progress in International Reading Literacy Study, PIRLS, som ble gjennomført i 35 land i 2001. Denne oppgaven tar kun for seg én datafil med observasjoner fra ett land, Norge, hvor det finnes informasjon som omfatter 3459 individer. Gjennom en prestasjonstest hvor elevene blir testet i sine leseferdigheter har vi data for hvordan fjerdeklasse elever gjør det på skolen i lesing. I tillegg til lesetesten har foreldrene til barnet, læreren og rektor av skolen fått utdelt spørreskjemaer med spørsmål om henholdsvis familie- og elevbakgrunn, lærer- og klassekarakteristika og skolekarakteristika. Dette gir oss informasjon som er nyttig når det skal undersøkes hvilke faktorer som kan være prestasjonsfremmende for barn.

Aktuell datafil:

- data2901-Norway.dta

Variabel	Forklaring
read	Test score/leseprestasjoner
kinderg_att	Barnehagedeltakelse, dummyvariabel med verdien 1 hvis de har gått i barnehage, 0 ellers
par_edu	Foreldrenes utdannings grad, kategorivariabel
par_not_born	Innvandrerbakgrunn, dummyvariabel med verdien 1 dersom foreldre ikke er født i Norge, 0 ellers.
income	Familieinntekt i dollar, kategorivariabel
girl	Elevens kjønn, dummyvariabel med verdien 1 dersom eleven er jente, 0 ellers.

Tabell 3.1: Definisjon av variabler

3.3 Deskriptiv statistikk for avhengig variabel

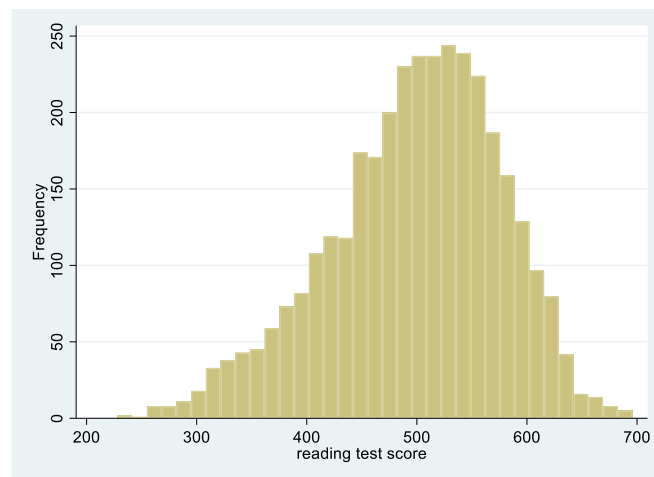
I denne oppgaven er leseprestasjoner, *read*, den avhengige variabelen. Deskriptiv statistikk for leseprestasjonene til alle fjerdeklassinger i Norge som har deltatt i undersøkelsen presenteres i tabell 3.2.

read	
Gjennomsnitt	498.2563
Standardavvik	78.36616
Min	228.0606
Max	695.8717
Antall observasjoner	3459

Tabell 3.2: Deskriptiv statistikk for avhengig variabel

Vi leser fra tabellen at det laveste testresultatet er 228.06 poeng og det høyeste er 695,87 poeng, hvor vi har et gjennomsnitt på 498.26 poeng. Standardavviket er på 78.37 poeng, som forteller oss hvor mye det gjennomsnittlige avviket fra 498,26 er. Antall observasjoner i undersøkelsen er på 3459.

Spredningen i observasjonene vises i et histogram presentert i figur 3.1. Denne figuren viser fordelingen av leseprestasjonene i Norge som vi ser er sentrert rundt toppene og dermed er en tilnærmet normalfordelt kurve.



Figur 3.1: Histogram som viser fordeling av leseferdigheter i Norge

3.4 Deskriptiv statistikk for interessevariabel

Vi ønsker å undersøke betydningen av å gå i barnehage når det kommer til hvordan elevene gjør det på skolen. Interessevariabelen er derfor *kinderg_att*, hvor tilhørende deskriptiv statistikk vises i tabell 3.3.

kinderg_att	
Gjennomsnitt	0.8604
Standardavvik	0.3467
Min	0
Max	1
Antall observasjoner	3137

Tabell 3.3: Deskriptiv statistikk for interessevariabel

Variabelen *kinderg_att* er en dummyvariabel som betyr at dersom eleven har gått i barnehage tar variabelen verdien 1, og 0 dersom eleven ikke har det. Den deskriptive statistikken må derfor tolkes litt annerledes enn for en kontinuerlig variabel, slik som *read*. Standardavviket forteller oss noe om variasjonen i variabelen da dette er gjennomsnittlig avvik fra gjennomsnittet. Vi ser fra tabellen at 86,04 % av de 3137 observasjonene har svart at barna gikk i barnehage før de begynte på skolen.

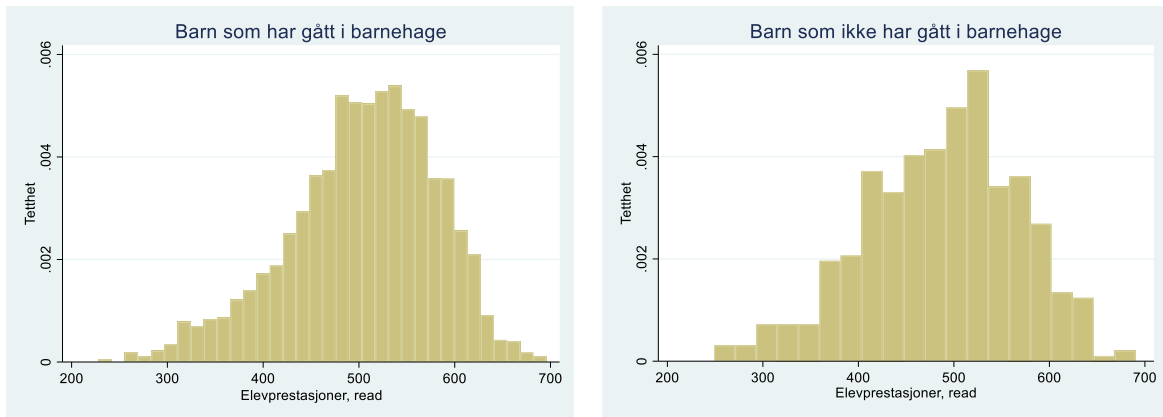
Siden interessevariabelen, *kinderg_att*, er en dummyvariabel vil informasjonen om korrelasjonen mellom *kinderg_att* og den avhengige variabelen, *read*, i et såkalt kryssplot, gi lite relevant informasjon. Det som istedenfor er mer interessant å se på er informasjonen om hvilke gjennomsnittsnivåer for elevprestasjoner det er innenfor de to gruppene, altså de som har gått i barnehage og ikke. Da kan vi få et innblikk i om det finnes forskjeller i prestasjon mellom elever som har gått i barnehage og de som ikke har gjort det.

	read	
	Gått i barnehage	Ikke gått i barnehage
Gjennomsnitt	504.0436	486.9031
Standardavvik	77.1062	81.2111
Min	228.0606	248.8043
Max	695.8717	690.4763
Antall observasjoner	2699	438

Tabell 3.4: Deskriptiv statistikk for avhengig variabel – gått i barnehage versus ikke gått i barnehage

Ut ifra denne tabellen kan vi observere et høyere gjennomsnittlig resultat for barn som har gått i barnehage enn for de som ikke har det. Likevel må det understrekes at antall observasjoner av barn som har gått i barnehage er over 6 ganger så mange som observasjoner av barn som ikke har det. Dette kan medføre skjevhet i resultatene og er noe som må tas høyde for når det konkluderes. I tillegg kan vi ikke med sikkerhet tolke dette som en kausal sammenheng, altså sammenhengen mellom en årsak og virkning, fordi det er mange flere faktorer som potensielt spiller inn på elevprestasjonene i tillegg til betydningen av å gå i barnehage.

Det kan også være informativt å se på denne oversikten presentert som et histogram. Figur 3.2 viser en fordeling av prestasjonsresultatene i de to gruppene. Vi ser at resultatene for både de elevene som har gått i barnehage og de som ikke har gjort det er normalfordelt og at fordelingen er nokså like i de to gruppene.



Figur 3.2: Histogram som viser fordelingen av prestasjonsresultater i de to gruppene.

3.5 Deskriptiv statistikk for kontrollvariablene

Det er sannsynlig at det er andre variabler enn interessevariabelen som også påvirker elevprestasjonene, og har derfor valgt å inkludere noen kontrollvariabler. Disse blir i kapittel 2.2 representert ved bokstavene P,F og S i skoleproduktfunksjonen, likning (0).

Inkluderte kontrollvariabler *par_edu*, *par_not_born*, *income* og *girl*. Forklaringer av variablene er presentert i tabell 3.1 tidligere i oppgaven.

Fordi det gir et mer meningsfullt svar i regresjonsanalysen senere endres kategorivariablene til dummyvariabler.

For kategorivariabelen om foreldrenes utdanning, *par_edu*, har jeg laget en dummyvariabel med navnet *Dedu_higher*. Dersom dummyvariabelen har en verdi lik 1 har forelder høyere utdanning, kategori 1 eller 2 i tabell 3.5, mens dersom verdien er 0 har forelder lavere eller ingen utdanning, kategori 3, 4 eller 5 i tabell 3.5. De uten utdanning er blitt slått sammen med de med lav utdanning fordi det er relativt få personer uten utdanning som har svart på undersøkelsen.

	par_edu	Dedu_higher
	Grad	Kategorivariabel
University degree	1	1
Post secondary (not university)	2	1
Upper secondary	3	0
Lower secondary	4	0
Not completed lower secondary	5	0

Tabell 3.5: Videre variabelforklaring for foreldrenes høyeste fullførte utdanning med omgjøring til dummyvariabel.

For kategorivariabelen om familiens inntekt, *income*, har jeg valgt å lage en dummy, *Dincome30000plus*, som tar verdien 1 for de som har en inntekt på mer enn \$30 000 eller mer, kategori 3,4,5 og 6 i tabell 3.6, og verdien 0 ellers.

income		<i>Dincome30000plus</i>
US \$	Kategorivariabel	Dummyvariabel
Mindre enn \$20,000	1	0
\$20,000 - \$29,999	2	0
\$30,000 - \$39,999	3	1
\$40,000 - \$49,999	4	1
\$50,000 - \$59,999	5	1
\$60,000 eller mer	6	1

Tabell 3.6: Videre variabelforklaring for årlig familieinntekt med omgjøring til dummyvariabel

I tabell 3.7 vises det deskriptiv statistikk for kontrollvariablene hvor vi ser at det er litt flere som har høyere utdanning enn lavere, at det er relativt få barn som har utenlandsfødte foreldre, at 83,7 % har en familie inntekt på 30 000 dollar eller mer og det er noe færre jenter.

	Gjennomsnitt	Standardavvik	Min	Max	Antall observasjoner
Dedu_higher	0.5429	0.4982	0	1	3098
par_not_born	0.0584	0.2345	0	1	3374
Dincome30000plus	0.8372	0.3692	0	1	3459
girl	0.4810	0.4997	0	1	3401

Tabell 3.7: Deskriptiv statistikk for kontrollvariablene.

For å nyansere informasjonen kan det være relevant å se på deskriptiv statistikk for gruppen som har gått i barnehage og de som ikke har gjort det hver for seg. Dette presenteres i tabell 3.8.

	Gått i barnehage			Ikke gått i barnehage		
	Gj. snitt	Std. avvik	Antall obs.	Gj. snitt	Std. avvik	Antall obs.
Dedu_higher	0.5707	0.4951	2660	0.3762	0.4850	428
par_not_born	0.0487	0.2154	2646	0.0631	0.2434	428
Dincome30000plus	0.8251	0.3799	2699	0.7968	0.4028	438
girl	0.4844	0.4999	2659	0.5255	0.4999	432

Tabell 3.8: Deskriptiv statistikk for kontrollvariablene – gått i barnehage versus ikke gått i barnehage.

Tabellen viser at elevene som har gått i barnehage oftere har foreldre med høyere utdanning. Den viser også det er noen flere barn med foreldre som er født i utlandet som ikke har gått i barnehage. Kjønn og familieinntekt ser ut til å være relativt jevnt fordelt i de to gruppene.

3.6 Korrelasjonsmatrise for datamaterialet

En korrelasjonsmatrise sier noe om potensialet for materialet og undersøker om variablene er lineært uavhengig av hverandre. Tallene i matrisen viser korrelasjonen mellom to variabler, altså R , men sier ikke hvor mye en variabel forklares av den andres variasjon. Dersom vi skal se på forklaringskraften til variablene må vi kvadrere R .

Dersom to variabler er perfekt korrelert vil R være lik 1, og høy korrelasjon vises ved at tallet nærmer seg 1. Vi ser i tabellen at ingen av variablene er svært høyt korrelert. Korrelasjonen mellom de med høy inntekt og høy utdanning er på 0.22, noe som vil si at de med høy utdanning oftere også har høy inntekt. Vi ser også en slik tendens mellom variablene *read* og *Dedu_higher*, noe som betyr at de trekker i samme retning. Når det er motsatt fortegn på korrelasjonen betyr det at de trekker i forskjellig retning.

	<i>read</i>	<i>kinderg_att</i>	<i>Dedu_higher</i>	<i>par_not_born</i>	<i>Dincome300~s</i>	<i>girl</i>
<i>read</i>	1.0000					
<i>kinderg_att</i>	0.0784	1.0000				
<i>Dedu_higher</i>	0.2916	0.1370	1.0000			
<i>par_not_born</i>	-0.1255	-0.0093	-0.0333	1.0000		
<i>Dincome300~s</i>	0.1218	0.0302	0.2204	-0.1418	1.0000	
<i>girl</i>	0.1284	-0.0314	0.0097	-0.0084	0.0144	1.0000

Tabell 3.9: Korrelasjonsmatrise for datamateriale

Et potensielt problem er dersom det er høy korrelasjon mellom interessevariabel og kontrollvariabel, for eksempel mellom variabel for om barnet har gått i barnehage (*kinderg_att*) og foreldrenes utdanning (*Dedu_higher*). Dersom det er høy korrelasjon mellom disse kan vi få et multikollinearitetsproblem, noe som betyr at variablene er til en viss grad gjensidig overflødige.

Et annet potensielt problem er utelatt variabelproblem. Dette skjer dersom man utelater en viktig variabel som viser seg å ha betydning for resultatet, eller gir et annet svar dersom man tar høyde for faktoren. Det oppstår med andre ord skjevhet i resultatene. Noen av forklaringsvariablene (kontrollvariablene eller interessevariabelen) vil få en større eller mindre forklaringskraft enn det den i virkeligheten har dersom vi hadde tatt hensyn til alle de relevante variablene. Gjenstående variabler tar effekten fra de utelatte og vi får upresis estimering. Dette kan for eksempel være å utelate variabelen om foreldrenes utdanning når man ser på betydningen av barnehage på leseprestasjoner. Dersom det viser seg at det er en positiv sammenheng mellom det å gå i barnehage og få gode testresultater på lesetesten kan det gi et feilaktig svar dersom vi utelater variabelen om foreldrenes utdanning. Dette er fordi det er grunn til å anta at utdanning på foresatte har en påvirkning på barnets utvikling, og det er også grunn til å tro at foreldre med høy utdanning ofte benytter barnehage til barna sine.

Dersom barn med foreldre som har lav utdanning sjeldnere har barn som går i barnehage og leseferdighetene varierer systematisk mellom barn som har foreldre med høy utdanning og foreldre som har lav utdanning har vi et potensielt utelatt variabelproblem.

3.7 Oppsummering

Datamaterialet som oppgaven baserer seg på er presentert i tillegg til definisjoner av relevante variabler. Kapitlet tar også for seg deskriptivstatistikk for avhengig variabel, interessevariabel og kontrollvariabler hvor det også skilles mellom barn som har gått i barnehage og de som ikke har. Vi har sett at det er en stor andel av respondentene som har gått i barnehage og at dette kan medføre skjevhet. I tillegg observeres det forskjeller på de to gruppene spesielt med tanke på foreldres utdanning og fødested. Til slutt ser vi på korrelasjonen mellom variablene, men observerer ingen urovekkende høye korrelasjoner og presenter potensielle problemer ved utelatt variabel og multikollinearitet.

4 Økonometrisk modell

4.1 Innledning

For å kunne svare på problemstillingen trenger vi å vite litt mer enn kun deskriptiv statistikk for de ulike variablene. I dette kapitlet forklares fremgangsmåten til den økonometriske modellen med relevante variabler som kan brukes til å teste ulike hypoteser.

4.2 Valg av funksjonsform

For å gjøre det litt enklere å skrive opp likningene uten at de blir for lange, blir variabelnavnet til interessevariabelen og kontrollvariablene forkortet. De forkortede variabelnavnene er presentert i tabell 4.1 og vil heretter bli brukt videre i oppgaven. Dummyvariabelen for kjønnnet på eleven har fremdeles variabelnavnet *girl*.

Variabel	Forkortet navn
kinderg_att	kindg
Dedu_higher	edu
par_not_born	born
Dincome30000plus	inc

Tabell 4.1: Forkortelse av variabelnavn.

Velger i denne oppgaven å ta utgangspunkt i en lineær approksimasjon av skoleproduktfunksjonen (0) som er beskrevet i kapitlet om teoretisk rammeverk. Senere i oppgaven skal det diskuteres utvidelser som tar for seg interaksjonsledd mellom et par av variablene.

Følgende likning settes opp når det tas utgangspunkt i at elevprestasjoner, *read*, er avhengig av om barnet har gått i barnehage eller ikke:

$$read = f(kindg, x) \quad (1)$$

Variabelen *kindg* er en dummyvariabel som forteller oss om barnet har gått i barnehage eller ikke. Denne likningen kan beskrive i hvilken grad barnehagen påvirker elevprestasjoner på skolen dersom variabelen blir inkludert korrekt.

Likning 2 er en enkel likning som beskriver elevprestasjoner som en funksjon av *kindg*.

$$read_i = \alpha + \delta * kindg_i + \beta * X_i + \varepsilon_i \quad (2)$$

Her er den avhengige variabelen *read* et mål på elevprestasjoner, *kindg* er en uavhengig dummyvariabel som forteller oss om barnet har gått i barnehage eller ikke, og *X* (linjevektor) er et sett med kontrollvariabler (*edu*, *born*, *inc*, *girl*) med en tilhørende koeffisientvektor β . Koeffisienten uttrykker relativ endring, sett alt annet likt.

- i* En tellevariabel som passer på at likningen beskriver alle individene i utvalget.
- α Likningens konstantledd.
- $\delta * kindg_i$ δ er en koeffisient der størrelsen på den beskriver betydningen av å ha gått i barnehage, og er lik for alle individene. Dersom δ går mot null er betydningen av å ha gått i barnehage liten og elevprestasjonene til eleven vil kun være avhengig av α , og dermed likt for både elever som har gått i, og ikke gått i barnehage. Betydningen av å ha gått i barnehage har derimot mye å si dersom δ er stor, og elevprestasjonene kan derfor variere i stor grad mellom elevene som har gått i barnehage og ikke.
- ε_i Likningens restledd som representerer andre variabler som påvirker resultatet, men som ikke er blant de spesifiserte.

Når vi inkluderer kontrollvariablene får vi:

$$read_i = \alpha + \delta * kindg_i + \beta_1 * edu_i + \beta_2 * born_i + \beta_3 * inc_i + \beta_4 * girl_i + \varepsilon_i \quad (3)$$

- $\beta_1 * edu_i$ *edu* varierer fra individ til individ, mens β_1 er dens tilhørende koeffisient som forteller hvor stor betydningen av å ha foreldre med høyere utdanning er på leseprestasjonene til eleven.
- $\beta_2 * born_i$ *born* varierer fra individ til individ, mens β_2 er dens tilhørende koeffisient som beskriver hvor stor betydningen av å ha foreldre med innvandrersstatus er på leseprestasjonene til eleven.

$\beta_3 * inc_i$ *inc* varierer fra individ til individ, mens β_3 er dens tilhørende koeffisient som forteller hvor stor betydningen av å ha en familieinntekt over 30 000 dollar er på leseprestasjonene til eleven.

$\beta_4 * girl_i$ *girl* varierer fra individ til individ, mens β_4 er dens tilhørende koeffisient som forteller hvor stor betydningen av å være jente er på leseprestasjonene til eleven.

Ønsker også å undersøke om en potensiell effekt av barnehagedeltakelse avhenger av utdanningen eller innvandrerstatusen til foreldrene, og velger derfor å inkludere noen interaksjonsledd.

Utvider først modellen med et interaksjonsledd mellom *kindg* og *edu* for å undersøke om betydningen effekten barnehage avhenger av foreldrenes utdanning.

$$read_i = \alpha + \delta * kindg_i + \beta_1 * edu_i + \rho_1 * kindg_i * edu_i + \beta_2 * born_i + \beta_3 * inc_i + \beta_4 * girl_i + \varepsilon_i \quad (4)$$

$\rho_1 * kindg_i * edu_i$ Dette er et interaksjonsledd hvor både *kindg* og *edu* enten har verdien 0 eller 1 og varierer fra individ. ρ_1 er korrelasjonskoeffisienten. Dersom $\rho_1 > 0$ betyr det at barnehagedeltakelse vil ha større effekt på leseprestasjonene blant barn som har foreldre med høy utdanning. Er $\rho_1 \approx 0$ forteller det at barnehagedeltakelse vil ha like mye å si uavhengig av foreldrenes utdanning, mens får vi $\rho_1 < 0$ har barnehagedeltakelse mindre å si for barn som har foreldre med høy utdanning enn de som har foreldre med lavere utdanning.

Det er også interessant å se på om foreldrenes innvandrerstatus har betydning for effekten av å ha gått i barnehage. Inkluderer derfor et interaksjonsledd mellom variabelen *kindg* og *born* i modell (5).

$$read_i = \alpha + \delta * kindg_i + \beta_1 * edu_i + \beta_2 * born_i + \rho_2 * kindg_i * born_i + \beta_3 * inc_i + \beta_4 * girl_i + \varepsilon_i \quad (5)$$

$\rho_2 * kindg_i * born_i$ Dette er et interaksjonsledd hvor både *kindg* og *born* enten har verdien 0 eller 1 og varierer mellom individer. ρ_2 er korrelasjonskoeffisienten. Dersom $\rho_2 > 0$ betyr det barnehagedeltakelse vil ha større effekt på leseprestasjonene blant barn med foreldre som har innvandrerbakgrunn.

Er $\rho_1 \approx 0$ forteller det at barnehagedeltakelse vil ha like mye å si både for barn med foreldre med innvandrerbakgrunn og de som ikke har det, mens får vi $\rho_1 < 0$ har barnehagedeltakelse mindre å si for barn som har foreldre med innvandrerbakgrunn enn de som ikke har det.

Tar utgangspunkt i likning (4) for å bedre beskrive mekanismen til interaksjonsleddet, og deriverer denne med hensyn på edu i likning (6).

$$\frac{\partial read_i}{\partial edu_i} = \beta_1 + \rho_1 * kindg_i \quad (6)$$

Her ser vi at størrelsen på ρ_1 beskriver hvor mye $kindg$ har å si på hvor mye elevprestasjoner avhenger av barnehagedeltakelse.

Det samme gjelder for interaksjonsleddet med foreldrenes innvandrerstatus, hvor ρ_2 forteller oss hvor mye $kindg$ har å si på hvor mye elevprestasjoner avhenger av barnehagedeltakelse når vi deriverer likning (4) med hensyn på $born$ i likning (7).

$$\frac{\partial read_i}{\partial born_i} = \beta_2 + \rho_2 * kindg_i \quad (7)$$

I likning (3), (4) og (5) forteller δ , β og ρ oss betydningen av variabelen som koeffisienten står foran. En koeffisient med verdi nærmere null reduserer betydningen av tilhørende variabel.

4.3 Empirisk strategi og estimeringsmetode

Hovedinteressen med denne oppgaven er å undersøke hvordan barnehagedeltakelse påvirker leseprestasjonene til elever i fjerde klasse. Dette gjøres ved å estimere δ i likning (2) og ved å teste hypoteser om den. For å kunne gjøre dette har vi inkludert en rekke kontrollvariabler i likning (3) som var representert ved X i (2) for å se hvordan estimatet δ endres og foreta hypotesetester når kontrollvariablene inkluderes. I tillegg ønsker vi å undersøke om effekten av barnehage avhenger av foreldrenes bakgrunn når det kommer til innvandrerstatus og grad av utdanning ved å inkludere interaksjonsledd.

For å finne sammenhengen mellom den avhengige variabelen, interessevariabelen og kontrollvariablene benytter vi estimeringsmetoden OLS. Dette er en minste kvadrats estimeringsmetode som finner løsningen der forskjellen mellom observasjonene og løsningen

er minst mulig, altså den sammenhengen som minimerer variansen. Når vi benytter denne metoden må vi foreta noen antakelser om observasjonene og datamaterialet.

Vi antar at sammenhengen mellom den avhengige variabelen og de uavhengige variablene er lineær og at modellen er additiv, altså at det ikke er statistisk interaksjon mellom variablene.

I tillegg må vi ha noen forutsetninger om restleddet, ε , for å kunne stole på de svarene vi får av regresjonen. Den første forutsetningen er at vi har en normalfordelt ε , noe som betyr at vi kan benytte t- og f- tester til å teste hypoteser om parametere. Forventningen til ε må være lik 0, noe som er tilfellet da det er inkludert et konstantledd α i modellen som tar opp variansen som ikke kan forklares av de andre forklaringsvariablene. Forklaringsvariablene kan heller ikke være korrelert med restleddet. Oppfylles disse betingelsene kan vil det si at estimatorene er forventningsrette, altså at den ikke er systematisk skjev. Modellen er også variansminimal dersom det ikke finnes estimatorene med lavere varians.

Vi skal i resultatdelen gjennomføre t-tester av koeffisientene. Når vi skal teste parametere må vi ha en alltid begynne med en nullhypotese i tillegg til en alternativ hypotese. Denne testen gjennomføres i dataprogrammet «stata» og gir oss en p-verdi. Denne verdien forteller oss om vi kan avfeie nullhypotesen ved at vi sammenligner den med signifikansnivået. Dersom p-verdien er lavere enn signifikansnivået kan nullhypotesen forkastes.

Det er vanlig å benytte et 95 % konfidensintervall, noe som betyr at vi har et areal på 5 % rundt regresjonslinja hvor det meste av dataen vår må være innenfor for at vi kan forkaste nullhypotesen ved et 5 % signifikansnivå.

R^2 er det som kalles for determinasjonskoeffisienten og er et mål på «føyning». Dette tallet sier noe om forklaringskraften til modellen, altså hvor mye som kan forklares av de variablene som er tatt med i modellen. R^2 er 1 minus residualvariasjonens andel av total variasjon, hvor det resterende er andelen som forklares av andre faktorer og tas opp i residualen, restleddet ε . Dersom R^2 er lik 1 betyr det at modellen forklarer all variasjon i den avhengige variabelen og dermed at alle variasjoner ligger på regresjonslinja. Ingenting av variasjonen blir forklart av modellen dersom R^2 er lik 0.

4.4 Oppsummering

I dette kapitlet ble valg av funksjonsform presentert der det tas utgangspunkt i skoleproduktfunksjonen (0), basert på tidligere litteratur. Det tar også for seg de ulike modellene som skal analyseres sammen med notasjonsforklaringer og hvordan vi tolker parameterne. Til slutt presenteres forutsetninger for OLS-estimatoren og forklaringer på hypotesetesting og forklaringskraften til modellen.

5 Empiriske resultater

5.1 Innledning

I dette kapittelet skal de empiriske resultatene presenteres og analyseres som følge av OLS-regresjonen. Det blir også gjennomført en rekke hypotesetester og til slutt en diskusjon av resultatene.

5.2 Empiriske hovedresultater

Resultatene fra likning (3) presenteres her sammen med en enkel regresjon uten kontrollvariablene som vi kaller likning (3.1) for å sammenligne.

$$read_i = \alpha + \delta * kindg_i + \varepsilon_i \quad (3.1)$$

$$read_i = \alpha + \delta * kindg_i + \beta_1 * edu_i + \beta_2 * born_i + \beta_3 * inc_i + \beta_4 * girl_i + \varepsilon_i \quad (3)$$

VARIABLES	(3.1) m0 read	(3) m1 read
kindg	17.141 (4.002)	9.570 (3.874)
edu		42.017 (2.752)
born		-39.253 (6.257)
inc		8.810 (3.599)
girl		19.374 (2.651)
Constant	486.903 (3.712)	457.442 (4.784)
Observations	3,137	2,998
R-squared	0.006	0.118

Standard errors in parentheses

Tabell 5.1: Presentasjon av OLS-regresjon med og uten kontrollvariabler.

Resultatene fra hovedproblemstillingen til oppgaven er presentert i tabell 5.1 med hvor det er kjørt en OLS-regresjon med og uten kontrollvariabler for å sammenligne. m0 presenterer resultatene for modellen uten kontrollvariablene. Her avhenger leseprestasjonene kun av om eleven har gått i barnehager eller ikke, og vi leser av tabellen at δ er 17.141. Dette kan tolkes slik at en elev som har gått i barnehage vil få 17.141 poeng mer sammenlignet med en elev

som ikke har det, noe som var forventet. Samtidig ser vi at forklaringskraften til modellen er betraktelig lav, hvor $R^2 = 0.006$, noe som kan skyldes få forklaringsvariabler i modellen og at det er sannsynlig at det er flere faktorer som har betydning for testresultatet.

m1 presenterer modellen med interessevariabel og alle kontrollvariablene. Det vi da ser er at δ reduseres til 9.570 samtidig som vi får en økning i forklaringskraften til $R^2 = 0.118$. Modellen har fremdeles en relativt lav forklaringskraft noe som tyder på at det er mange andre faktorer som også har effekt på resultatet, men som ikke tas opp i modellen.

Videre leser vi at bakgrunnen til foreldrene har relativt stor effekt på resultatet i denne modellen, men i motsatt retning, noe som var forventet. Barn med foreldre med høy utdanning scorer i snitt 42.017 poeng bedre enn barn som ikke har det. Innvandrerstatusen har negativt fortegn og barn med foreldre som ikke er født i Norge gjør det i snitt 39.253 poeng dårligere sammenlignet med et barn hvor foreldrene er født i Norge i denne modellen. Leser også at både familiens inntekt og det å være jente har en positiv effekt på testresultatet, noe som også var forventet selv om betydningen av jente muligens var høyere enn antatt.

For å kunne stole på noen av resultatene i modellen tester vi om effekten til hver av forklaringsvariablene er statistisk signifikante. Vi benytter en t-test hvor nullhypotesen, H_0 , er at variabelen ikke har betydning for leseferdighetene til fjerdeklasseelever, mens alternativ hypotesen, H_A , at det har en betydning.

$$H_0: \delta = 0$$

$$H_A: \delta \neq 0$$

Kjører testen i «stata» og får en p-verdi lik 0.0135 som betyr at vi kan avvise nullhypotesen på et 5 % signifikansnivå om at det ikke finnes noen sammenheng mellom barnehagedeltakelse og leseprestasjoner.

Gjennomfører den samme testen for hver av beta-koeffisientene, hvor

$$H_0: \beta_i = 0$$

$$H_A: \beta_i \neq 0$$

og får at p-verdien til både *edu*, *born* og *girl* er lik 0.000, noe som betyr at vi kan forkaste nullhypotesen med disse variablene ved alle signifikansnivå og det utelukkende finnes en sammenheng mellom disse og hvordan eleven presterer på skolen. Variabelen *inc* har en p-verdi på 0.0144 og vi kan derfor forkaste nullhypotesen ved et 5 % signifikansnivå.

5.3 Empiriske resultater av tillegsspørsmål

I tillegg ønsket vi å undersøke hvorvidt effekten av barnehagedeltakelse avhenger av foreldrenes bakgrunn når det kommer til grad av utdanning og innvandrerstatus. Dette gjøres ved å inkludere interaksjonsledd i modellen. Resultatene fra likning (4) og (5) presenteres i tabell 5.2 sammen med hovedresultatene fra tabell 5.1.

$$read_i = \alpha + \delta * kindg_i + \beta_1 * edu_i + \rho_1 * kindg_i * edu_i + \beta_2 * born_i + \beta_3 * inc_i + \beta_4 * girl_i + \varepsilon_i \quad (4)$$

$$read_i = \alpha + \delta * kindg_i + \beta_1 * edu_i + \beta_2 * born_i + \rho_2 * kindg_i * born_i + \beta_3 * inc_i + \beta_4 * girl_i + \varepsilon_i \quad (5)$$

VARIABLES	(3.1) m0 read	(3) m1 read	(4) m2 read	(5) m3 read
kindg	17.141 (4.002)	9.570 (3.874)	10.298 (5.015)	7.876 (3.978)
edu		42.017 (2.752)	43.580 (7.374)	42.113 (2.752)
kindg_edu			-1.806 (7.905)	
born		-39.253 (6.257)	-39.278 (6.259)	-66.409 (15.905)
inc		8.810 (3.599)	8.818 (3.600)	8.842 (3.598)
girl		19.374 (2.651)	19.398 (2.654)	19.277 (2.651)
kindg_born				32.039 (17.253)
Constant	486.903 (3.712)	457.442 (4.784)	456.837 (5.470)	458.872 (4.843)
Observations	3,137	2,998	2,998	2,998
R-squared	0.006	0.118	0.118	0.119

Standard errors in parentheses

Tabell 5.2: Presentasjon av modeller inkludert OLS-regresjon med interaksjonsledd.

Vi tar først for oss OLS-regresjonen m2 som inkluderer et interaksjonsledd mellom *kindg* og *edu*, representert ved *kindg_edu*. De andre variablene tolkes likt som tidligere. Koeffisienten til interaksjonsleddet kan ikke ses på uten å se på de andre variablene som forklarer den sammen effekten. Vi ser først på endringer fra (3) til (4), hvor δ har økt med $10.298 - 9.570 = 0.728$ noe som betyr at introduksjonen av interaksjonsleddet førte til en økt betydning av det å gå i barnehage. Dette gjelder imidlertid kun for de barna som har foreldre uten høy utdanning. Vi ser at $\rho_1 = -1.806$ noe som indikerer at barnehagedeltakelse har mindre betydning for barn med foreldre med høy utdanning.

Gjennomfører en t-test i stata for å undersøke om interaksjonsleddet er statistisk signifikant forskjellig fra null og har hypotesene:

$$H_0: \rho_1 = 0$$

$$H_A: \rho_1 \neq 0$$

Får p-verdien 0.8085 og vi kan ikke forkaste nullhypotesen ved noen signifikansnivåer. Dette betyr at det er veldig lite sannsynlig at utdanningen til foreldrene har noen betydning for effekten av å gå i barnehage. Effekten er med andre ord lik for barn med høyt utdannede foreldre og barn med lavere utdannede foreldre.

Deretter tar vi for oss OLS-regresjonen m3 som inkluderer et interaksjonsledd mellom *kindg* og *born*, representert ved *kindg_born*. De andre variablene tolkes også her likt som tidligere og koeffisienten til interaksjonsleddet kan ikke ses på uten å se på de andre variablene som forklarer den sammen effekten. Ser nå på endringer i estimatene fra likning (3) til (5) og leser at δ har en endring på $7.876 - 9.570 = -1.694$. Dette betyr at betydningen av å gå i barnehage er mindre for de som har foreldre som er født i Norge. Når vi leser av koeffisienten til interaksjonsleddet finner vi $\rho_2 = 32.039$ noe som tyder på en økt effekt av å gå i barnehage dersom du har foreldre med innvandrerbakgrunn.

Gjennomfører også her en t-test i stata for å undersøke om interaksjonsleddet er statistisk signifikant forskjellig fra null og har hypotesene:

$$H_0: \rho_2 = 0$$

$$H_A: \rho_2 \neq 0$$

Når vi kjører testen i stata gir den oss p-verdien 0.0634 og vi kan forkaste hypotesen på et 10 % signifikansnivå. Dette betyr at vi kan være relativt sikre på at innvandrerstatusen til foreldrene har betydning for effekten av barnehagedeltakelse og at det har en større effekt for barn med foreldre som ikke er født i Norge.

5.4 Diskusjon

Resultatene av den empiriske analysen tyder på at det finnes en positiv sammenheng mellom barnehagedeltakelse og gode resultater på leseprestasjonstesten. Denne effekten reduseres når kontrollvariablene tas med, men forklaringskraften styrkes noe som tyder på at det er et mer riktig resultat. Som nevnt tidligere i oppgaven er det langt flere som har svart at de har gått i barnehage noe som tyder på høy barnehageoppslutning i Norge, men kan også gi et upresist svar med tanke på utvalget.

Gjennom inkludering av kontrollvariablene bekrefte teorien om at det er mange faktorer som spiller inn på elevens leseprestasjoner da en t-test av effekten var statistisk signifikant. Vi ser særlig at foreldre med høyere utdanning har en stor positiv betydning for leseferdighetene til barna, og at foreldre som er født utenfor Norge påvirker i nesten like stor grad, men i motsatt retning. Ser også at kjønn har stor betydning, hvor jenter scorer bedre enn gutter, mens inntekten til familien har omtrent like stor betydning som barnehage.

I tilleggsproblemstillingen ønsket vi å undersøke om effekten av barnehage avhenger av foreldrenes bakgrunn med tanke på innvandrerstatus og utdanning. Her må det også poengteres at variabelen *born* som er brukt for å identifisere innvandrere tar for seg alle elever som har foreldre som ikke er født i Norge, altså faller de som er født i nordiske land også under denne kategorien. I tillegg er hvorvidt barna er født i Norge ikke spesifisert i denne oppgaven da fokuset har vært på foreldrenes bakgrunn.

I resultatene av tilleggsproblemstillingen fant vi en signifikant sammenheng mellom effekten av barnehagedeltakelse og det å ha foreldre som ikke er født i Norge. Dette tyder på at viktigheten av å gå i barnehage gjelder spesielt for disse barna. En grunn til dette kan være at de språklige ferdighetene man tilegner seg i barnehagen har stor betydning for videre læring og at barn med innvandrerforeldre får utviklet disse i høyere grad dersom de går i barnehage. Det var derimot ingen signifikant sammenheng mellom effekten av barnehagedeltakelse og foreldre med høy utdanning. Dette indikerer at barnet har like stor effekt av å gå i barnehage uavhengig av foreldrenes grad av utdanning.

5.5 Oppsummering

I dette kapittelet har resultatene for regresjonsanalysen blitt presentert og til slutt diskutert. Det tar for seg hovedresultatene om betydningen av barnehage, men også resultatene for tilleggsproblemstillingen som omhandler hvorvidt effekten av barnehagedeltakelse avhenger av foreldrenes bakgrunn. Det gjennomføres også en rekke hypotesetester for å undersøke om resultatene er signifikante.

6 Konklusjon

Vi startet med problemstillingen om hvordan barnehagedeltakelse påvirker leseprestasjoner av å gå i barnehage i Norge. En modell med kun barnehagedeltakelse alene som uavhengig variabel og deretter en utvidelse med 4 kontrollvariabler om elevens familiekarakteristika da det er grunn til å tro at det er flere faktorer som har betydning for elevens leseferdigheter. Denne utvidelsen styrket modellens forklaringskraft, R^2 , og både interessevariabelen og alle kontrollvariablene var signifikante når t-tester ble gjennomført.

Som forventet så vi en positiv sammenheng mellom barnehagedeltakelse og leseprestasjoner, noe som tyder på at norske barnehager både er viktig for individets utvikling, men også samfunnsøkonomisk sett da dette bidrar til å øke ferdighetsnivået i befolkningen. På grunn av de store forskjellene av effekten av barnehagedeltakelse mellom barn med innvandrerforeldre og ikke, tyder det på at språklige ferdigheter er en viktig faktor for videre læring.

7 Videre arbeid

I denne analysen er ikke skole- eller klassekarakteristika tatt med, noe som kan føre til et utelatt variabelproblem fordi det muligens ville påvirke størrelsen til barnehagevariabelen sin koeffisient enten i den ene eller andre retningen. Det kan også tenkes at det finnes uobserverbare karakteristika som korrelerer med effekten av å gå i barnehage og dermed ikke tas opp.

Det er kun tatt for seg data fra Norge i denne oppgave og det diskuteres derfor ikke hvordan resultatene ville variert mellom land. Et interessant spørsmål er om den positive effekten av barnehage kun gjelder i det norske systemet eller om det også er tilfelle i andre land.

En utvidelse av analysen som kan være relevant er å skille mellom barnehager med tanke på dens størrelse, kompetansen til arbeiderne og sammensetningen av barn. Dette kan bidra til å undersøke hvilke faktorer innad i barnehagene som fører til gode resultater på skolen.

Referanser

- [1] Regjeringen.no (2019), Kvalitet i barnehagen. Hentet fra <https://www.regjeringen.no/no/tema/familie-og-barn/barnehager/artikler/kvalitet-i-barnehagen/id2612951/> [6. mai 2020]
- [2] Korsvold, T. (2017), barnehage-historie. Hentet fra https://snl.no/barnehage_-_historie [6. april 2020].
- [3] Korsvold, T. (2017), barnehage. Hentet fra <https://snl.no/barnehage> [6.april 2020]
- [4] Hanushek, E. A. (2020): Education production functions. I Bradley, S. og Green, C. (red): Economics of Education, 2nd Edition, London: Academic Press, 161-170.
<http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202020%20Education%20Production%20Functions.pdf>
- [5] Bonesrønning, H (2004), Utforming av utdanningspolitikken.
<https://samfunnsokonomene.no/wp-content/uploads/2019/05/Trykkutgave-4-2012.pdf>
- Thomas, R. L. (2005): Using statistics in economics. McGraw Hill Education

Appendix

Stata-resultater:

`. reg read kindg`

Source	SS	df	MS	Number of obs	=	3,137
Model	110715.914	1	110715.914	F(1, 3135)	=	18.34
Residual	18922731.6	3,135	6035.95903	Prob > F	=	0.0000
Total	19033447.5	3,136	6069.33912	R-squared	=	0.0058
				Adj R-squared	=	0.0055
				Root MSE	=	77.691

read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
kindg	17.14051	4.002137	4.28	0.000	9.293439 24.98758
_cons	486.9031	3.71224	131.16	0.000	479.6245 494.1818

Figur A.1: Stata-resultater for likning (3.1)

`. reg read kindg inc edu born girl`

Source	SS	df	MS	Number of obs	=	2,998
Model	2095737.66	5	419147.531	F(5, 2992)	=	79.70
Residual	15735496.9	2,992	5259.19013	Prob > F	=	0.0000
Total	17831234.5	2,997	5949.69454	R-squared	=	0.1175
				Adj R-squared	=	0.1161
				Root MSE	=	72.52

read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
kindg	9.570392	3.873859	2.47	0.014	1.974696 17.16609
inc	8.809813	3.599351	2.45	0.014	1.75236 15.86727
edu	42.01665	2.752163	15.27	0.000	36.62033 47.41297
born	-39.25318	6.257435	-6.27	0.000	-51.5225 -26.98387
girl	19.37399	2.651218	7.31	0.000	14.17559 24.57238
_cons	457.4423	4.783815	95.62	0.000	448.0624 466.8222

Figur A.2: Stata-resultater for likning (3)

```
. reg read kindg inc edu kindg_edu born girl
```

Source	SS	df	MS	Number of obs	=	2,998
Model	2096012.38	6	349335.397	F(6, 2991)	=	66.40
Residual	15735222.2	2,991	5260.85662	Prob > F	=	0.0000
				R-squared	=	0.1175
				Adj R-squared	=	0.1158
Total	17831234.5	2,997	5949.69454	Root MSE	=	72.532

read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
kindg	10.29817	5.015409	2.05	0.040	.4641748 20.13217
inc	8.81816	3.600106	2.45	0.014	1.759225 15.8771
edu	43.57984	7.373566	5.91	0.000	29.12207 58.03762
kindg_edu	-1.806487	7.905187	-0.23	0.819	-17.30664 13.69367
born	-39.27835	6.259395	-6.28	0.000	-51.5515 -27.00519
girl	19.39819	2.653751	7.31	0.000	14.19482 24.60155
_cons	456.8365	5.469775	83.52	0.000	446.1116 467.5614

Figur A.3: Stata-resultater for likning (4)

```
. reg read kindg edu born kindg_born inc girl
```

Source	SS	df	MS	Number of obs	=	2,998
Model	2113858.34	6	352309.724	F(6, 2991)	=	67.04
Residual	15717376.2	2,991	5254.89007	Prob > F	=	0.0000
				R-squared	=	0.1185
				Adj R-squared	=	0.1168
Total	17831234.5	2,997	5949.69454	Root MSE	=	72.491

read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
kindg	7.875605	3.978374	1.98	0.048	.0749775 15.67623
edu	42.11321	2.751529	15.31	0.000	36.71813 47.50829
born	-66.40944	15.90544	-4.18	0.000	-97.59616 -35.22273
kindg_born	32.03917	17.25344	1.86	0.063	-1.790645 65.86899
inc	8.841576	3.59792	2.46	0.014	1.786928 15.89622
girl	19.27654	2.650653	7.27	0.000	14.07926 24.47383
_cons	458.872	4.843444	94.74	0.000	449.3752 468.3688

Figur A.4: Stata-resultater for likning (5)

