

Randi Stølan & Bettina Wold

Effekt av barnehage på elevprestasjoner

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2020

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

Randi Stølan & Bettina Wold

Effekt av barnehage på elevprestasjoner

Bacheloroppgave i Samfunnsøkonomi
Veileder: Bjarne Strøm
Mai 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Sammendrag

Det er bred konsensus blant forskere og fagmiljø om at barnehagedeltakelse gir positive utslag på elevprestasjoner. Hensikten med denne oppgaven er å analysere barnehageeffekten, og se om den endres ved å ta høyde for andre faktorer som kan ha betydning for elevprestasjonene. Analysen baserer seg på datamateriale fra PIRLS 2001, en internasjonal undersøkelse om lesing og lesevaner blant 4. klassinger, hvor det er brukt datasett fra Norge og Sverige.

Sverige plasserer seg øverst blant alle landene som deltok i leseprøven, mens Norge havner rett under gjennomsnittet. I tillegg til å se på den generelle effekten av barnehage, vil vi undersøke om barnehageeffekten kan være utslagsgivende for forskjellene i leseferdigheter mellom Norge og Sverige. Resultatene fra analysen viser en klar sammenheng mellom deltakelse i barnehage og elevprestasjoner, men forskjellene i ferdighetsnivået mellom landene som kan forklares av barnehagedeltakelse er uklar. Vi ser derimot en sammenheng mellom sosioøkonomisk bakgrunn og effekten av barnehage.

Innhold

1. Innledning	3
2. Teoretisk rammeverk og tidligere litteratur	3
2.1 Teoretisk rammeverk.....	3
2.2 Litteratur og tidligere funn	4
3. Økonometrisk tilrettelegging.....	6
3.1 Innledning.....	6
3.2 Regresjonsanalyse med to variabler	6
3.2.1 <i>Populasjons- og utvalgsregresjonsmodellen</i>	6
3.2.2 <i>Minste kvadraters metode (OLS)</i>	6
3.2.3 <i>Egenskaper ved OLS-estimatorene</i>	7
3.2.4 <i>Hypotesetesting</i>	8
3.3 Multipel regresjon	9
4. Datamaterialet	10
4.1 Innledning.....	10
4.2 Om datamaterialet	10
4.2.3 <i>Definisjoner og forklaringer av variabler</i>	10
4.2.4 <i>Deskriptiv statistikk</i>	12
4.2.5 <i>Svakheter ved datasettet</i>	14
5. Regresjonsanalyse	15
5.1 Innledning.....	15
5.2 Regresjonsmodeller	15
5.2.1 <i>Ulike sett av kontrollvariabler</i>	15
5.2.2 <i>Forskjeller mellom Norge og Sverige</i>	17
5.2.3 <i>Betydningen av barnehage i Norge og Sverige</i>	18
5.2.4 <i>Tilleggsproblemstilling</i>	20
5.3 Multikollinearitet.....	21
5.4 Heteroskedastisitet.....	22
5.5 Tolkning av resultater.....	22
6. Oppsummering	24
6.1 Konklusjon	24
6.2 Videre studier	25
7 Referanseliste	26
Appendiks.....	27

Effekt av barnehage på elevprestasjoner

1. Innledning

Elevprestasjoner varierer mellom land. Vi ønsker i denne oppgaven å se på elevprestasjoner i Norge og Sverige, to land med tilsynelatende like grunnforutsetninger, men som ifølge PIRLS 2001 scorer ulikt på en omfattende leseprøve. Vi vil først og fremst se på hvordan deltakelse i barnehage påvirker leseferdigheter generelt, fordi det regnes som en arena for tidlig læring. Helt konkret blir problemstillingen følgende:

Hvilken betydning har barnehagedeltakelse for leseferdigheter? Kan forskjellene i leseferdigheter i Norge og Sverige begrunnes med at de har ulik effekt av å gå i barnehage?

Vår hypotese er at barnehagedeltakelse har positiv effekt på leseferdigheter. Vi ser på to land med vidt forskjellige resultater, derfor vil det også være interessant å se om barnehageeffekten er utslagsgivende for ulikhetene.

Til slutt vil vi undersøke om barn fra forskjellige hjem har ulik effekt av barnehage. Dette blir en tilleggsproblemstilling, og vil ikke bli like omfattende som den øvrige analysen. Vi har likevel valgt å inkludere det i oppgaven, fordi vi i undersøkelsene av hovedproblemstillingen fant flere tidligere studier som tyder på at barn fra hjem med lav sosioøkonomisk status har bedre effekt av barnehage.

2. Teoretisk rammeverk og tidligere litteratur

2.1 Teoretisk rammeverk

I denne delen skal vi etablere en skoleproduktfunksjon, som kan brukes til å presentere resultater i skolen. Produktfunksjonen er en statistisk modell for ferdigheter (T) som vil avhenge av en rekke faktorer. Her presenteres en enkel form av produktfunksjonen, som senere benyttes i den økonometriske analysen.

$$T = f(S, F, P) \quad (2.1)$$

S - her inngår skolefaktorer som klassestørrelse og lærerkarakteristikka

F - individspesifikke faktorer som familie- og elevkarakteristikka

P - medelevkarakteristikka

Medelevkarakteristikka tar hensyn til effekter på ferdighetene av å være i en gruppe med jevnaldrende. Gruppene inneholder ofte hierarkier og distinkte atferdsmønstre, noe som kan påvirke ferdighetene hos det enkelte individet. Slike effekter kan for eksempel komme frem gjennom variabler som omhandler forhold og holdninger i hjemmet. Om de fleste barna kommer fra velstående hjem med mange bøker, kan det tenkes at det danner seg en kultur for å tilegne seg kunnskap blant gruppene.

Opgaven tar utgangspunkt i PIRLS 2001, en omfattende undersøkelse med 35 deltagende land (Solheim, G. & Tønnesen, E., 2003). Undersøkelsen består av en leseprøve for 4. og 5. klassinger, og i tillegg til leseprøven ble det utformet spørreskjema til både elever, foresatte og ansatte i skolen som omhandlet alle forhold som inngår i skoleproduktfunksjonen.

2.2 Litteratur og tidligere funn

Det er bred forskning på utdanningspolitikk av økonomer, men også fra andre fagfelt. Barnehage synes å være en viktig arena for tidlig læring, da dette gir et fortrinn for senere læring (Havnes, T & Mogstad, M. 2009). Effekten av å gå i barnehage er ikke bare et skandinavisk fenomen, men noe man kan se flere steder i verden. I 1927 ble det utført undersøkelser på en barneskole, av Faye Risser og Harry Elder ved universitetet i Chicago, der de sammenlignet barn som hadde deltatt i barnehage med de som ikke hadde det. Funnene tyder på at leseferdighetene var bedre for samtlige fra 1. til 5. klasse som hadde deltatt i barnehage (Risser, F & Elder, H. 1927).

Tidligere studier tyder også på at tidlig innsats har større effekt enn kompenserende tiltak senere, og at man har positiv kognitiv effekt av å gå i barnehage i Norge (NOU 2019: 3). Utredningen er relativt ny, men legger blant andre den tidligere nobelprisvinneren og amerikanske forskeren Joseph Heckman til grunn. Heckman er anerkjent for sin forskning på feltet som omhandler tidlig læring, og omtaler det å gå i barnehage som en

multiplikatoreffekt. Han argumenterer for at satsing på tidlig læring har stor betydning for videre læring på skolen (Heckman, J. 2000). Heckman sine studier viser at det også er ulikt ferdighetsnivå blant de som har gått i barnehage, og mener at forholdene hjemme spiller en avgjørende rolle for denne avstanden (Heckman, J. 2006).

Sosiologen James Coleman mente også at forholdene i hjemmet var av stor betydning, og var godt kjent for spissformuleringen «it's all in the family». Colemans forskning tyder på at S-vektoren i skoleproduktfunksjonen ikke har vist signifikant effekt. Blant økonomer er det ulike meninger om denne effekten, til tross for bred politisk enighet om at mindre klasser indikerer kvalitet på undervisningen (Bonesrønning, H. 2004).

De ovennevnte forskerne er enig om at barnehagedeltakelse har positiv effekt på leseferdigheter, men strides om i hvor stor grad de ulike vektorene i skoleproduktfunksjonen påvirker leseferdighetene. Skal man tro Coleman har F-vektoren stor betydning. Vi vil derfor, som en tilleggsproblemstilling, se på om barn med ulik familiebakgrunn har forskjellig effekt ved å gå i barnehage.

Økonomen Elizabeth Dhuey har forsket på effekten av barnehagedeltakelse i USA, og fant i sine studier at barn fra ikke-engelsktalende innvandrerfamilier og barn fra familier med lav sosioøkonomisk standard hadde større avkastning av barnehagedeltakelse. Det trekkes blant annet frem at barn fra latinamerikanske familier med tilgang til barnehage hadde 17 prosent mindre sannsynlighet for å havne under nivået for sitt årskull. I tillegg fant Dhuey at latinamerikanske barn med tilgang på barnehage fikk 5 prosent høyere lønn som voksen (Dhuey, E. 2007, s. 2).

Det er også blitt gjennomført lignende studier i Norge. Ifølge Havnes, T. og Mogstad, M. (2009) hadde barn fra familier med lavt utdanningsnivå mest nytte av barnehage. Aller størst var effekten for de barna som hadde mødre med lav utdanning (s. 29). Mye tyder på at sosioøkonomisk bakgrunn spiller en betydelig rolle for hvilken effekt barn har av barnehagedeltakelse.

3. Økonometrisk tilrettelegging

3.1 Innledning

I denne delen vil vi gi en innføring i metoden som brukes i den senere analysen, samt forklare de hovedbegreper og hovedmekanismer bak sammenhengene som skal undersøkes.

3.2 Regresjonsanalyse med to variabler

3.2.1 Populasjons- og utvalgsregresjonsmodellen

Vi antar for enkelthets skyld at vi har en produktfunksjon med kun en forklaringsvariabel x . Vi vil da få følgende likning:

$$read_i = \alpha + \beta_1 x + \varepsilon_i \quad (3.1)$$

Read er forventede leseferdigheter gitt forklaringsvariabelen x . α og β_1 er ukjente populasjonsparametre, ε er restleddet og i er et telleledd, som i dette tilfellet står for antall land.

Likning (3.1) kalles en populasjonsregresjonslikning. De ukjente populasjonsparametrene viser henholdsvis konstantleddet og helningen på en rett linje, da (3.1) representerer et lineært forhold. Restleddet viser innflytelsen på leseprestasjon av alle andre variabler enn den relevante forklaringsvariabelen. Fordi populasjonsregresjonslikningen er ukjent, bruker vi data for den avhengige variabelen og x for å anslå α og β . Estimaten i den predikerte modellen, utvalgsregresjonsmodellen, blir a og b .

Forskjellen på verdien av leseferdigheter og predikert verdi for leseferdigheter kalles en residual, og er gitt ved e . Det er den estimerte verdien for restleddet ε . I et diagram vil man få ulike punkter som reflekterer ulike observasjoner, og vårt problem blir å beregne verdier på a og b slik at man får en rett linje som passer best mulig til observasjonene.

3.2.2 Minste kvadraters metode (OLS)

For å beregne a og b kan man bruke minste kvadraters metode, også kjent som OLS-regresjon. Man vil velge a og b slik at residualene blir minst mulig, og forskjellen mellom

faktiske variabler og predikerte variabler blir små.

Sammenhengen mellom read og x fra (3.1) viser at når x øker med en enhet, gir det en økning i leseferdighet med β enheter. Vi får da en predikert økning i leseferdighet med b enheter.

Hvor mye av variasjonen i leseferdigheter som faktisk kan forklares av vår modell, vil avhenge av hvor stor del av variasjonen som er forklart og uforklart.

$$R^2 = \frac{\text{Total variasjon} - \text{uforklart variasjon}}{\text{Total variasjon}} = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST}$$

Der $0 \leq R^2 \leq 1$

R^2 kalles en determinasjonskoeffisient, og er et mål på føyning. Dersom verdien er 1, forklarer modellen all variasjon i leseferdigheter. Ved verdi 0 forklarer modellen ingenting av variasjonen. Jo lavere uforklart variasjon er i forhold til total variasjon, desto høyere forklaringskraft vil modellen ha. Kvadratrotten av R^2 , R , viser korrelasjonen, og forteller hvor sterk samvariasjon det er mellom variabler.

3.2.3 Egenskaper ved OLS-estimatorene

Vi ønsker at en estimator skal være forventningsrett og effisient. Med forventningsrett menes det at estimatoren ikke skal avvike systematisk fra den sanne parameterverdien. Med effisient menes det at den skal gi mest mulig presis informasjon om de ukjente parameterne. Hvor presis en estimator er måles ved variansen. En forventningsrett estimator er effisient hvis det ikke finnes andre estimatorer med lavere varians.

For at OLS-estimatorene skal være forventningsrette må en rekke forutsetninger være oppfylt. Man må ha en klassisk situasjon med ikke-stokastiske x'er som har faste verdier i repeterte utvalg. Verdiene må også være slik at når antall observasjoner øker og går mot uendelig, går variansen mot en fast konstant (Thomas, 2005, s. 357). Det betyr at variansen ikke øker uendelig mye. I en slik situasjon får man følgende forutsetninger om restleddet:

i) $E(\varepsilon_i) = 0$

ii) $Var(\varepsilon_i) = \sigma^2 < \infty$

iii) $Cov(\varepsilon_i, \varepsilon_j) = 0$ for alle $i \neq j$

iv) Hver ε_i er normalfordelt

Når forutsetning *ii*) holder, altså at variansen til restleddene er konstant og ikke varierer med størrelsen på forklaringsvariablene, vil restleddene være homoskedastiske (Thomas, 2005, s. 360). Dersom man har stor grad av barnehagedeltakelse i Norge og Sverige, og det i tillegg er stor variasjon i utvalget i Norge og Sverige, kan det være en utfordring for denne forutsetningen. Det kan da tenkes at det landet med størst utvalg har større variasjon i leseferdigheter, og dermed også høyere varians i restleddet. Dette vil vi se nærmere på i regresjonsanalysen i del 5.

Under klassiske forutsetninger leses a og b standard normalfordelt. Når man kjenner standardavviket kan man teste hypoteser om β ved hjelp av standard normalfordelingen. Det er normalt at variansen og dermed standardavviket er ukjent, og standardavviket må estimeres. Forventningsrett estimator for standardavviket er:

$$S = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Det estimerte standardavviket til b -verdiene blir dermed forholdet mellom S og summen av avvik fra faktiske verdier fra variablenes respektive utvalgsgjennomsnitt.

3.2.4 Hypotesetesting

Kjenner nå estimert verdi for b og standardavviket, og kan teste hypoteser om β ved hjelp av standard normalfordelingen.

Eksempel: Test $H_0: \beta = \beta_0$ mot $H_A: \beta > \beta_0$

Dersom $\beta = \beta_0$ betyr det at leseferdigheter ikke er påvirket av x . Dersom $\beta > \beta_0$ vil vi forkaste nullhypotesen. Beslutningsregelen blir da at man forkaster H_0 dersom

$$TS = \frac{b - \beta_0}{S_b} > t_{\alpha, n-2}$$

Der TS er testobservator, t er kritisk verdi, α i dette tilfellet er signifikansnivået og $n-2$ er antall frihetsgrader, der n er antall observasjoner.

3.3 Multipel regresjon

Ved multipel regresjon har man samme oppsett som ved likning (3.1), men man inkluderer flere kontrollvariabler i modellen for å se om det er andre parametre som kan forklare variasjonen i leseferdigheter:

$$read_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon_i \quad (3.2)$$

I dette tilfellet vil man få enda en forutsetning om forklaringsvariablene, som sier at det ikke skal være et lineært forhold mellom utvalgsverdiene til to eller flere forklaringsvariabler.

Dette er for å unngå perfekt multikollinearitet (Thomas, 2005, s. 394). Ved perfekt multikollinearitet har man komplett usikkerhet når det gjelder populasjonsverdiene for β . Problemet oppstår når to eller flere variabler har korrelasjon på 1, og forklarer akkurat det samme. Stata vil utelate eventuelle variabler som har dette problemet, men det kan likevel oppstå høy multikollinearitet.

Grunnen til at vi velger å inkludere kontrollvariabler, er at det er viktig å se på flere enn to variabler for å begrense feilspesifikasjon. Feilspesifikasjon betyr at man utelater relevante forklaringsvariabler, som vil gi skjeve OLS-estimer dersom noen av de utelatte variablene er korrelert med de noen av de inkluderte. Fordi vi ikke har med alle variablene fra datasettet i analysen, kan det tenkes at noen av de inkluderte variablene tar opp effekter fra utelatte variabler.

Ved multipel regresjon kan man utføre hypotesetesting som ved regresjon med to variabler, men antall frihetsgrader avhenger av antall forklaringsvariabler. I vårt tilfelle er det så mange respondenter at vi ser på uendelig antall frihetsgrader i t-fordelingen i alle våre hypotesetester.

Ved multipel regresjon kan man også bruke F-test. Her ser man på endring i uforklart variasjon, SSR, når man legger på restriksjoner i modellen. Nullhypotesen kan forkastes dersom følgende kriterium er oppfylt: $TS > F_{\alpha, (h, n-k)}$, der

$$TS = \frac{R_u^2 - R_r^2/h}{1 - R_u^2/n - k}$$

R_u^2 og R_r^2 betegner forklaringskraft i hhv. den urestrikkerte og den restrikkerte modellen. Antall restriksjoner betegnes av h , og k er antall forklaringsvariabler.

Skoleproduktfunksjonen krever at vi kontrollerer for andre variabler enn hvorvidt respondentene har gått i barnehage eller ikke hvis vi skal få en troverdig modell. Det vil derfor være hensiktsmessig å bruke multipl regressjonsanalyse.

4. Datamaterialet

4.1 Innledning

I dette kapitlet presenteres datasettet som benyttes. Vi vil definere og forklare variablene som er brukt i analysen, og vise deskriptiv statistikk for de ulike variablene.

4.2 Om datamaterialet

Oppgaven tar utgangspunkt i PIRLS (Progress in International Reading Literacy Study) 2001, som ble presentert tidligere, og datasettet er basert på denne undersøkelsen. Alle respondentene fra Norge og Sverige gikk i 4. klasse da undersøkelsen ble gjennomført. Vi bruker dataprogrammet Stata til å utføre regressjonene.

PIRLS ble satt i gang av IEA (The International Association for the Evaluation of Educational Achievement). IEA var allerede i 1991 ansvarlige for den første internasjonale leserundersøkelsen som Norge deltok i, og har hatt en sentral rolle i utarbeiding av leseprøvene og spørreskjemaene for PIRLS (Solheim, G & Tønnesen, E. 2003).

4.2.3 Definisjoner og forklaringer av variabler

<i>read</i>	Kontinuerlig variabel som viser elevenes testscore.
<i>kinderg_att</i>	Dummyvariabel hvor verdi=1 dersom respondenten har gått i barnehage, 0 ellers.
<i>clsiz</i>	Kontinuerlig variabel som viser antall elever i klassen.
<i>clsiz_sq</i>	Variabel for klassestørrelse kvadrert.

<i>teacher_fem</i>	Dummyvariabel hvor verdi=1 dersom respondentens lærer er kvinne, 0 ellers.
<i>zage</i>	Kontinuerlig variabel for alder, standardisert.
<i>girl</i>	Dummyvariabel hvor verdi=1 hvis individet er jente, 0 ellers.
<i>par_not_born</i>	Dummyvariabel hvor verdi=1 hvis individet har foreldre som er født i utlandet, 0 ellers.
<i>early_ability_x</i>	Kategorivariabel omkodet til dummysett som viser respondentenes tidlig litterære ferdigheter.
<i>pct_disadv_x</i>	Kategorivariabel omkodet til dummysett som viser prosentandelen av elever som kommer fra økonomisk vanskeligstilte hjem.
<i>books100plus</i>	Kategorivariabel omkodet til dummyvariabel. Variabelen har verdi=1 dersom respondenten har 100 eller flere bøker i hjemmet, 0 ellers.
<i>norway</i>	Dummyvariabel med verdi=1 dersom respondenten kommer fra Norge, 0 ellers. Det vil si at Sverige er referansegruppe.
<i>kinderg_nor</i>	Samspillsvariabel mellom variablene <i>kinderg_att</i> og <i>norway</i> . Verdi=1 dersom respondenten både er fra Norge og har gått i barnehage. Verdi=0 dersom respondenten er fra Sverige, og dersom respondenten er fra Norge og ikke har gått i barnehage.
<i>books100_kinderg</i>	Samspillsvariabel mellom variablene <i>books100plus</i> og <i>kinderg_att</i> . Verdi=1 dersom respondenten har gått i barnehage og har 100 bøker eller flere i hjemmet. Verdi=0 dersom respondenten ikke har gått i barnehage, og dersom respondenten har gått i barnehage men ikke har 100 bøker eller flere i hjemmet.

For at det skal være mulig og hensiktsmessig å sammenligne kategorivariablene med resten av datasettet har de blitt omkodet til dummysett der en av dummiene blir brukt som referansekategori i analysen. *Early_ability* er en kategorivariabel hvor elevene svarer på om de hadde tidlig litterære ferdigheter. Variabelen har følgende verdier: ikke i det hele tatt=1, ikke veldig bra=2, moderat bra=3 og veldig bra=4. For å kunne tolke variabelen i analysen har vi generert variabelen *early_ability_x*, der x=1-4. *Early_ability_1* har verdien 1 for de som svarer «ikke i det hele tatt», og 0 ellers. Vi har brukt *early_ability_1* som referansekategori.

Kategorivariabelen *pct_disadv* oppgir prosentandelen skoleelever som kommer fra økonomisk vanskeligstilte hjem. Dummysettet er basert på følgende kategorier: 0-10%, 11-25%, 26-50% og 50% eller flere. Vi har valgt å bruke variabelen hvor 50% eller flere kommer fra økonomisk vanskeligstilte hjem som referansekategori.

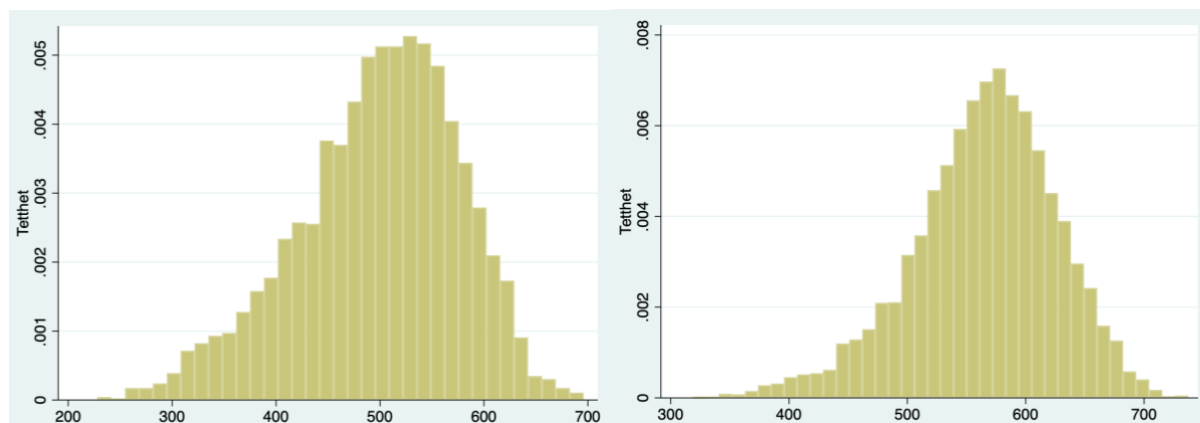
Books_home er også en opprinnelig kategorivariabel, men vi har valgt en mer restriktiv tilnærming hvor den er delt opp i 2 kategorier istedenfor 5, som den opprinnelig hadde. Vi så det hensiktsmessig å skille ved 100 bøker i hjemmet for å se på sosioøkonomiske forskjeller, og antar dermed at de som har flere enn 100 bøker i hjemmet kommer fra ressurssterke familier sammenlignet med de som ikke har det. Dette er en sterk forutsetning og kan kun brukes som et utgangspunkt for videre undersøkelser.

4.2.4 Deskriptiv statistikk

	<i>Begge</i>	<i>Norge</i>	<i>Sverige</i>
<i>Observasjoner</i>	10658	3459	7199
<i>Gjennomsnittsverdi</i>	543.14	498.26	564.70
<i>Standardavvik</i>	(74.161)	(78.366)	(66.313)
<i>Minimumsverdi</i>	228.06	228.06	318.68
<i>Maksverdi</i>	737.33	695.87	737.35

Tabell 1: Deskriptiv statistikk for avhengig variabel.

Fra den deskriptive statistikken ser vi at svenskene har både høyere gjennomsnittsverdi og høyere maks- og minimumsverdier på testscore. Standardavviket til Sverige er mindre enn for Norge, som tilsier at elevene fra Sverige har mindre spredning og holder et jevnere nivå.



Figur 1: Histogram for Norge og Sverige.

Toppnivået i histogrammene viser testscorene som er vanligst blant elevene. Denne toppen er høyere i Sverige, som vises til høyre, enn i Norge. Testscore i begge landene er konsentrert rundt toppene og kurvene kan dermed sies å være normalfordelt. I Norge er det noe større differanse mellom de som har gjennomsnittlig testscore og de som scorer under snittet.

	<i>Begge</i>	<i>Norge</i>	<i>Sverige</i>
<i>Gjennomsnitt</i>	0.92	0.86	0.95
<i>Standardavvik</i>	(0.275)	(0.346)	(0.227)

Tabell 2: Deskriptiv statistikk for interessevariabel.

Ser her at en høyere prosentandel fra den svenske befolkningen deltar i barnehage (95%), sammenlignet med Norge (86%).

	<i>Begge</i>		<i>Norge</i>		<i>Sverige</i>	
	Gj.snitt	Standardavvik	Gj.snitt	Standardavvik	Gj.snitt	Standardavvik
<i>clsiz</i>	24.71	(8.287)	20.96	(4.826)	26.59	(8.989)
<i>teacher_fem</i>	0.83	(0.376)	0.87	(0.336)	0.81	(0.393)
<i>age</i>	10.53	(0.507)	9.97	(0.332)	10.79	(0.332)
<i>girl</i>	0.49	(0.499)	0.48	(0.499)	0.49	(0.499)
<i>par_not_born</i>	0.10	(0.302)	0.06	(0.235)	0.12	(0.328)
<i>early_ability</i>	2.78	(0.944)	2.62	(0.947)	2.83	(0.935)
<i>pct_disadv</i>	1.60	(0.849)	1.18	(0.445)	1.79	(0.916)
<i>books100plus</i>	0.70	(0.456)	0.70	(0.457)	0.70	(0.456)

Tabell 3: Deskriptiv statistikk for kontrollvariabler.

Klassestørrelsene er mindre i Norge med 21 elever per klasse i snitt, mens Sverige har nærmere 27 elever per klasse. Både Norge og Sverige har en høy andel kvinnelige lærere, som begge er over 80 prosent. Alder er lavere i Norge enn Sverige, som er naturlig ettersom norske elever starter på grunnskolen når de er seks år, mens man starter det året man fyller syv i Sverige. Antall jenter i klassene i de to landene er tilnærmet likt og utgjør omtrent halvparten av utvalget.

En større andel av barna i Sverige har foreldre født i utlandet. Den deskriptive statistikken for *early_ability* og *pct_disadv* er målt ved kategorivariablene. Vi ser da at gjennomsnittet i Norge og Sverige ligger nærmere 3 ved *early_ability*, der 3 betyr at eleven hadde moderat gode ferdigheter. Ved *pct_disadv* ligger verdien på hhv. 1.18 og 1.79. Husker at 1 står for de skolene som hadde 0-10 prosent elever fra økonomisk vanskeligstilte hjem, og i Norge ligger gjennomsnittet rett over dette. I Sverige er gjennomsnittet nærmere at 11-25 prosent av elevene på skolen er fra økonomisk vanskeligstilte hjem.

Målt ved variabelen *books100plus* viser den deskriptive statistikken at Sverige og Norge har ca. like mange familier med 100 eller flere bøker i hjemmet, relativt sett.

4.2.5 Svakheter ved datasettet

Som nevnt starter grunnskoleelever i Norge som seksåringer og har derfor en noe lavere alder enn de svenske elevene. Vi har derfor standardisert age til zage, slik at vi kan sammenligne datasett med stor forskjell i aldersvariabelen på tvers av ulike geografiske områder. I tillegg kan det være det en svakhet at majoriteten av utvalget er svenske, da Sverige kan trekke gjennomsnittsverdiene opp eller ned med ekstremverdier.

En annen svakhet kan være såkalte missing values, som oppstår når respondentene i undersøkelsen ikke har svart på samtlige spørsmål. Når man omkoder kategorivariabler til dummier er det en tendens til at missing values blir med i variabelen under verdi 1. Ved for eksempel *books100plus*, kan forholdet i realiteten være litt under 70 prosent, noe som ikke kommer fram i den deskriptive statistikken. Avvikene er små, fordi de fleste respondentene har svart på kontrollvariablene som brukes i analysen.

5. Regresjonsanalyse

5.1 Innledning

$$read = f(kinderg_att, x) \quad (5.1)$$

Vi vil undersøke hvordan estimatet på *kinderg_att* endres og foreta hypotesetester når ulike sett av kontrollvariabler inkluderes i likning (5.1). Vi vil også se på betydningen av eventuelle ikke-lineariteter i effekten av testobservatoren ved å inkludere samspillsledd. Videre vil vi kvalitetssikre modellen ved å se om den oppfyller noen av forutsetningene som ble presentert i del 3, før vi til slutt tolker resultatene fra analysen. Vi vil ikke forklare variablene vi bruker i denne delen, da de allerede er forklart i del 4.2.3.

5.2 Regresjonsmodeller

5.2.1 Ulike sett av kontrollvariabler

Følgende modeller ser på ulike sett av kontrollvariabler for alle elementene i skoleproduktfunksjonen. Modell 1 ser på skolefaktorer og individkarakteristikk, mens modell 2 ser på individkarakteristikk samt familie- og medelevkarakteristikk.

Modell 1:

$$read_i = \alpha + \delta_1 * kinderg_att + \beta_1 * clsize + \beta_2 * clsize_sq + \delta_2 * teacher_fem + \beta_3 * zage + \delta_3 * girl + \varepsilon_i$$

Vi kan utføre en hypotesetest for å se om barnehagedeltakelse påvirker leseferdigheter. Siden vi har a priori grunn til å tro at barnehagedeltakelse har positiv effekt på leseferdigheter, kan vi formulere null- og alternativhypotesen på følgende måte:

$H_0: \delta_1 = 0$: Barnehagedeltakelse har ingen effekt på leseferdigheter

$H_A: \delta_1 > 0$: Barnehagedeltakelse har positiv effekt på leseferdigheter

Vi kan forkaste nullhypotesen om at barnehage ikke har effekt på leseferdigheter da $TS = 8,62 > t_{0.05} = 1,645$. Vi godtar indirekte alternativhypotesen om at barnehagedeltakelse har positiv effekt på leseferdigheter.

Modell 2:

$$read_i = \alpha + \delta_1 * kinderg_att + \delta_4 * par_not_born + \delta_{5-7} * early_ability_x + \delta_{8-10} * pct_disadv_x + \delta_{11} * books100plus + \varepsilon_i$$

Ved samme signifikansnivå og null- og alternativhypotese som ved modell 1 får vi at $TS = 8,93 > t_{0.05} = 1,645$. Vi kan forkaste nullhypotesen.

Alle de inkluderte variablene har positiv effekt på leseferdigheter, bortsett fra *par_not_born*. Kan utføre en hypotesetest her, for å se om denne effekten er signifikant. Siden man har grunn til å tro at de som har utenlandske foreldre scorer dårligere på leseferdigheter, kan man utføre en ensidig test med følgende hypoteser:

$H_0: \delta_4 = 0$: Foreldre født i utlandet har ingen effekt på leseferdigheter

$H_A: \delta_4 < 0$: Foreldre født i utlandet har negativ effekt på leseferdigheter

Dersom absoluttverdien av testobservatoren er større enn $t_{0.05} = 1,645$, kan vi forkaste nullhypotesen. $|TS| = 7,62$, og vi forkaster nullhypotesen.

Modell 3:

$$read_i = \alpha + \delta_1 * kinderg_att + \beta_1 * clsize + \beta_2 * clsize_sq + \delta_2 * teacher_fem + \beta_3 * zage + \delta_3 * girl + \delta_4 * par_not_born + \delta_{5-7} * early_ability_x + \delta_{8-10} * pct_disadv_x + \delta_{11} * books100plus + \varepsilon_i$$

Her har vi inkludert alle kontrollvariablene fra modell 1 og 2. Forklaringskraften, R^2 , øker betydelig, fra henholdsvis 16,6% og 20,7% til 30,9%. Barnehagedeltakelse får noe mindre betydning, men ved samme hypotesetest som under de foregående modellene får man at $TS = 6,08 > t_{0.05} = 1,645$ og man kan fortsatt forkaste nullhypotesen med 95% sikkerhet. Faktisk kan vi ved alle tre modellene forkaste nullhypotesen på alle signifikansnivå.

Det er verdt å merke seg den standardiserte aldersvariabelen, *zage*. Den sier at dersom alderen øker med ett aldersintervall, vil testscoren øke med ca. 21 poeng. Her har vi også valgt et 5% signifikansnivå, og formulert følgende hypoteser:

$H_0: \beta_3 = 0$: Økt alder har ingen effekt på leseferdigheter

$H_A: \beta_3 > 0$: Økt alder har positiv effekt på leseferdigheter

Får at $TS = 28,54 > t_{0,05} = 1,645$. Vi kan forkaste nullhypotesen både ved $t_{0,05}$ og alle andre signifikansnivå.

5.2.2 Forskjeller mellom Norge og Sverige

Modell 4:

$$\begin{aligned} read_i = & \alpha + \delta_1 * kinderg_{att} + \beta_1 * clsizex + \beta_2 * clsizex_{sq} + \delta_2 * teacher_{fem} + \\ & \beta_3 * zage + \delta_3 * girl + \delta_4 * par_{not_born} + \delta_{5-7} * early_{ability}_x + \\ & \delta_{8-10} * pct_{disadv}_x + \delta_{11} * books100plus + \delta_{12} * norway + \epsilon_i \end{aligned}$$

I modell 4 har vi introdusert dummyvariabelen *norway*. Vår interessevariabel, *kinderg_att*, viser fortsatt positiv effekt av å ha gått i barnehage. Med $TS = 4,46$ er den fortsatt signifikant ved $t_{0,05} = 1,645$. Konstantleddet på 441,1 viser testscoren til de svenske respondentene etter å ha kontrollert for de andre variablene samt *norway* i modell 4. Vi ser dermed at det å være fra Norge gir en negativ effekt på leseferdigheter, og de scorer i snitt 53 poeng dårligere på leseprøven enn hva svenske respondenter gjør.

$H_0: \delta_{12} = 0$: Hvorvidt respondenten er fra Norge har ingen effekt på leseferdigheter

$H_A: \delta_{12} < 0$: Hvorvidt respondenten er fra Norge har negativ effekt på leseferdigheter

Alternativhypotesen er formulert med bakgrunn i den deskriptive statistikken, som viser at Sverige har høyere gjennomsnittlig testscore. $|TS| = 12,21 > t_{0,05} = 1,645$. Vi ser at absoluttverdien til TS er langt høyere enn kritisk verdi, og forkaster nullhypotesen.

I forrige modell merket vi oss aldersvariabelen, som var statistisk signifikant. Ved samme hypotesetest på aldersvariabelen i modell 4 er den fortsatt signifikant, men har en betydelig lavere testobservator, på 4,51. Koeffisienten til *zage* forteller oss at testscoren øker mindre når respondenten sin alder øker med ett aldersintervall etter man har kontrollert for land.

5.2.3 Betydningen av barnehage i Norge og Sverige

Modell 5:

$$\begin{aligned} read_i = & \alpha + \delta_1 * kinderg_att + \beta_1 * clsize + \beta_2 * clsize_sq + \delta_2 * teacher_fem \\ & + \beta_3 * zage + \delta_3 * girl + \delta_4 * par_not_born + \delta_{5-7} * early_ability_x \\ & + \delta_{8-10} * pct_disadv_x + \delta_{11} * books100plus + \delta_{12} * norway + \rho_1 \\ & * kinderg_nor + \varepsilon_i \end{aligned}$$

For å se om betydningen av å ha gått i barnehage er ulik i Norge og Sverige, har vi dannet samspillsvariabelen *kinderg_nor*. Koeffisienten til *kinderg_att* vil nå vise effekten av å ha gått i barnehage for svenske individer. Konstantleddet viser leseferdighetene til svenske individer som ikke har gått i barnehage.

Vi tester fortsatt *kinderg_att* ved en ensidig hypotesetest, med samme hypoteser og signifikansnivå som i modell 1-4. Vi får da at $TS = 1,87$, dermed er den fortsatt signifikant. Betydningen er redusert fra tidligere modeller, og standardavviket har økt.

Når vi skal se på forskjeller i effekt ved å ha gått i barnehage, ser vi ingen åpenbar a priori grunn til at effekten skal være større i Norge enn i Sverige. Vi har derfor valgt en tosidig hypotesetest for variabelen *kinderg_nor*. Ved $t_{0,05}$ må $-1,96 > TS > 1,96$.

$H_0: \rho_1 = 0$: Ingen forskjell mellom landene i betydning av å ha gått i barnehage

$H_A: \rho_1 \neq 0$: Norge har bedre eller dårligere effekt av å gå i barnehage enn Sverige

Vi får at $TS = 1,76$. Absoluttverdien er dermed under kritisk verdi på 1,96, og vi kan ikke forkaste nullhypotesen med 95% sikkerhet.

Variablene *kinderg_att* og *kinderg_nor* har relativt høye standardavvik i modell 5. Vi utfører dermed en F-test for å se om vi kan si at de er signifikant kombinert. Får følgende null- og alternativhypotese:

$H_0: \delta_1 = \rho_1 = 0$

$H_A: \delta_1 = \rho_1 \neq 0$ Minst en av δ_1, ρ_1 er ikke lik 0

Vi får da $TS = 11,47 > F_{0,05, (2, 8046)} = 2,99$. Vi kan forkaste nullhypotesen og ser at de er signifikant kombinert. Det er likevel verdt å merke seg at den justerte determinasjonskoeffisienten øker lite fra modell 4 til modell 5. Variasjonen i leseferdigheter som kan tilskrives variabelen *kinderg_nor* kan dermed sies å være liten.

	(1) <i>read</i>	(2) <i>read</i>	(3) <i>read</i>	(4) <i>read</i>	(5) <i>read</i>
<i>kinderg_att</i>	21.95*** (2.545)	23.56*** (2.639)	15.26*** (2.511)	10.88*** (2.441)	6.517 (3.482)
<i>Kontrollert for skolefaktorer og individkarakteristikka</i>	Ja	Nei	Ja	Ja	Ja
<i>Kontrollert for individ-, familie- og medelevekarakteristikka.</i>	Nei	Ja	Ja	Ja	Ja
<i>books100plus</i>		28.30*** (1.564)	27.29*** (1.494)	26.75*** (1.448)	26.76*** (1.448)
<i>norway</i>				-53.06*** (2.318)	-60.89*** (4.980)
<i>kinderg_nor</i>					8.559 (4.875)
<i>cons</i>	457.9*** (6.548)	457.4*** (4.839)	407.7*** (7.401)	442.1*** (7.327)	445.9*** (7.635)
<i>N</i>	9105	8419	8063	8063	8063
<i>Adjusted R²</i>	0.1656	0.2064	0.3075	0.3497	0.3499

t statistics in parentheses

- $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tabell 4: Regresjonsmodell 1-5.

5.2.4 Tilleggsproblemstilling

Modell 6:

$$\begin{aligned} read_i = & \alpha + \delta_1 * kinderg_att + \beta_1 * clsize + \beta_2 * clsize_sq + \delta_2 * teacher_fem + \beta_3 \\ & * zage + \delta_3 * girl + \delta_4 * par_not_born + \delta_{5-7} * early_ability_x + \delta_{8-10} \\ & * pct_disadv_x + \delta_{11} * books100plus + \rho_2 * books100_kinderg + \varepsilon_i \end{aligned}$$

I tilleggsproblemstillingen har vi valgt å utelate variablene som ser spesifikt på forskjeller mellom landene, men har tatt høyde for de andre kontrollvariablene. Med bakgrunn i empiri presentert om sosioøkonomisk bakgrunn, har vi konstruert en samspillvariabel mellom *books100plus* og *kinderg_att*.

Variabelen *kinderg_att* viser nå hvilken betydning barnehagedeltakelse har for respondenter med færre enn 100 bøker i hjemmet. Testobservatoren til *kinderg_att* i modell 6 er 5,80 ved samme hypotesetest og signifikansnivå som ved modell 1-5. Vi tester om det å ha 100 bøker eller flere i hjemmet har positiv betydning for leseferdigheter:

$H_0: \delta_{11} = 0$: Det å ha 100 eller flere bøker i hjemmet har ingen effekt på leseferdigheter

$H_A: \delta_{11} > 0$: Det å ha 100 eller flere bøker i hjemmet har positiv effekt på leseferdigheter

Ved $t_{0,05}$ får vi $TS = 8$, og vi kan dermed forkaste nullhypotesen. Vi ser at koeffisienten til *books100_kinderg* viser -12,99. Det antyder at de som har 100 eller flere bøker i hjemmet har lavere effekt av barnehagedeltakelse enn de som ikke har det. Vi har tidligere etablert følgende null- og alternativhypotese, ut fra den empirien vi har tilgjengelig:

$H_0: \rho_2 = 0$: Hvorvidt respondentene kommer fra ressurssterke familier gir ingen forskjell i barnehageeffekt

$H_A: \rho_2 < 0$: Respondenter fra ressurssterke familier har lavere barnehageeffekt

$|TS| = 2,54 > t_{0,05} = 1,645$. Testobservatoren vi får er negativ, og ved å forkaste nullhypotesen kan vi indirekte godta at barn som har 100 bøker eller flere i hjemmet har lavere effekt av barnehage enn de som har færre enn 100 bøker i hjemmet.

	(3) <i>read</i>	(6) <i>read</i>
<i>kinderg_att</i>	15.26*** (2.511)	23.17*** (3.994)
<i>Kontrollert for skolefaktorer samt individ-, familie- og medelevkarakteristikk a</i>	Ja	Ja
<i>books100plus</i>		39.14*** (4.893)
<i>books100_kinderg</i>		-12.99* (5.105)
<i>cons</i>	407.7*** (7.401)	400.5*** (7.926)
<i>N</i>	8063	8063
<i>Adjusted R₂</i>	0.3075	0.3079

t statistics in parentheses

- $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tabell 5: Regresjonsmodell 3 og 6.

5.3 Multikollinearitet

Vi ser fra korrelasjonsmatrisen (vedlegg 3) at det er høy korrelasjon mellom *norway* og *kinderg_nor*. Det kan forklares ved at en stor andel av de norske respondentene har gått i barnehage, dermed vil variablene vise mye av det samme.

Det kan være en svakhet med datasettet, fordi det i tillegg til å gi høye standardavvik kan gi endringer i andre variabler enn de som er korrelerte. Samspillsvariabelen *kinderg_nor* kan da få store variasjoner i testobservator avhengig av hvilke variabler den kombineres med. I tillegg endres justert R_2 lite fra modell 4 til modell 5 ved å tilføre variabelen *kinderg_nor*. Vi kan utføre en F-test på variablene for å se om de er signifikant kombinert:

$$H_0: \delta_{12} = \rho_1 = 0$$

$$H_A: \delta_{12} = \rho_1 \neq 0 \text{ En av } \delta_{12}, \rho_1 \text{ er ikke lik } 0$$

Vi får $TS = 263,72 > F_{0,05, (2, 8046)} = 2,99$, og kan forkaste nullhypotesen.

Problemet med multikollinearitet gjør seg først gjeldende når R^2 er nært opp mot 1 i verdi. Vår høyeste determinasjonskoeffisient er på 0,35, dermed er vårt problem hovedsakelig at det er vanskelig å skille de korrelerte variablenes individuelle effekt på avhengig variabel.

5.4 Heteroskedastisitet

En annen forutsetning for å ha effisiente og forventningsrette OLS-estimatorer er homoskedastisitet. For å finne ut om forutsetningen oppfylles kan man teste for heteroskedastisitet, og en vanlig måte å gjøre dette på er ved bruk av Breusch-Pagan testen (Thomas, 2005, s. 483). Man undersøker om variansen i restleddene fra regresjonen er betinget av verdiene i de uavhengige variablene. Alle våre regresjonsmodeller tester positivt for heteroskedastisitet, som er et brudd på forutsetningen om homoskedastisitet i restleddet.

En måte å håndtere problemet på er å estimere regresjonene for robuste standardfeil, som kan gjøres i Stata. Vi finner at endringene i t-verdier er relativt små, samtidig som koeffisientene i alle modellene er uendret (vedlegg 2). Modellene fra analysen kan beholdes, da de ikke har tilstrekkelig grad av heteroskedastisitet til at det gjør betydelige utslag på standardfeilene.

5.5 Tolkning av resultater

Alle våre modeller viser at barnehagedeltakelse har positiv effekt på leseferdigheter. I modell 1-5 ser vi likevel at koeffisienten foran *kinderg_att* gradvis blir mindre når man tar høyde for flere faktorer fra skoleproduktfunksjonen. Det kan tolkes som at betydningen av å ha gått i barnehage reduseres når man ser på andre skole-, individ- og familiefaktorer; for eksempel økt alder, tidlig litterære ferdigheter og ressurssterke familier.

I modell 2 ser vi at de som har foreldre født i utlandet har dårligere leseferdigheter (vedlegg 1). Det kan være en indikator på at familieforhold er av betydning, dermed er det relevant å inkludere dette i skoleproduktfunksjonen. De som har foreldre født i utlandet ville kanskje hatt relativt større utbytte av barnehage.

Økt alder synes å ha positiv effekt på leseferdigheter, men koeffisienten reduseres fra modell 3 til 4 (vedlegg 1), når man inkluderer forskjellene mellom Norge og Sverige. Det kan indikere at Sverige har mindre effekt av økt alder, men fra den deskriptive statistikken ser vi at svenske respondenter i snitt er nesten ett år eldre enn norske respondenter. Det kan dermed tenkes at effekten av å bli eldre på leseferdigheter blir mindre jo eldre man blir, ut ifra det aldersintervallet vi ser på.

Modell 4 viser også at man har store forskjeller mellom Norge og Sverige i leseferdigheter. Norge har dårligere testscore enn Sverige, men begge landene har positiv effekt av barnehagedeltakelse. Når vi innfører variabelen *kinderg_nor* ser vi at den ikke er signifikant på et 5%-nivå. På et 10%-nivå ville den vært signifikant, og vi utelukker derfor ikke at norske respondenter kan ha større effekt av barnehagedeltakelse enn de svenske. Effekten må studeres videre for å kunne gi et mer solid svar, og oppgaven kan ikke konkludere med verken det ene eller det andre når det gjelder ulikhet i effekt av barnehagedeltakelse mellom landene. Det er viktig å skille mellom korrelasjon og kausalitet, og selv om variablene korrelerer i vår analyse kan vi ikke påstå at det er en kausal årsakssammenheng.

I modell 5 ser vi at betydningen av barnehagedeltakelse ikke er like stor når vi har kontrollert for land og skilt mellom effekten av barnehagedeltakelse i de to landene. I modell 6, derimot, får barnehage igjen økt betydning når man fjerner skillet mellom landene og ser på samspillet som representerer barn fra ressurssterke familier sin effekt av barnehagedeltakelse. Her er resultatet at respondenter fra ressurssterke familier har signifikant dårligere effekt av barnehage enn referansegruppen. Modellen får likevel lavere forklaringskraft, og skillet mellom land er kanskje derfor av større betydning. Med så store forskjeller i leseferdigheter mellom Norge og Sverige, kan det tenkes at mye av forklaringskraften i modell 4 og 5 kan tilskrives skillet mellom landene.

Enkelte av kontrollvariablene i analysen blir ikke alltid signifikante, men vi har valgt å beholde dem i modellen da vårt hovedproblem er å se på effekt av barnehagedeltakelse.

6. Oppsummering

6.1 Konklusjon

I denne oppgaven har vi sett på om barnehagedeltakelse har ulik effekt på leseferdigheter i Norge og Sverige, med utgangspunkt i datasett fra PIRLS 2001. Vi har sett på tidligere forskning som bakgrunn for oppgaven, hvor vi fant at det i stor grad argumenteres for de positive effektene av barnehage og tidlig læring, blant annet av Heckmann. I arbeidet med tidligere litteratur så vi at flere forskere hevder effekten av barnehage er større for enkelte grupper i samfunnet, som ble bakgrunn for vår tilleggsproblemstilling (Dhuey, 2007).

Vi har tatt utgangspunkt i skoleproduktfunksjonen, og sett på variabler for både individ-, familie- og medelevkarakteristikk. Analysen startet med en restriktiv modell som kun så på skolefaktorer og elevkarakteristikk, før vi utvidet med flere kontrollvariabler og samspillsledd for å se på forskjeller mellom land og forskjeller mellom elever fra ulike hjem.

Hovedproblemstillingen, som var om barnehagedeltakelse har positiv effekt på leseferdigheter eller ikke, besvares gjennomgående i regresjonsanalysen ved at barnehagedeltakelse fører til høyere testscore. Vi ser videre at effekten av å gå i barnehage blir mindre når vi kontrollerer for andre faktorer som påvirker leseferdigheter hos 4. klassinger, og det kan dermed sies at barnehagedeltakelse er en av mange faktorer som påvirker leseferdigheter. Vi får ikke et klart svar på om barnehageeffekt er utslagsgivende for forskjeller i elevprestasjoner mellom landene.

Tilleggsproblemstillingen belyses ved å se på om barn med 100 bøker eller flere har lavere effekt av barnehage, og funnene i oppgaven tyder på at elever med lavere sosioøkonomisk status har større effekt av barnehage enn de som kommer fra ressurssterke hjem.

Resultatene fra analysen må tolkes med forsiktighet. Tilleggsproblemstillingen er vag, da den kun er undersøkt ved bruk av én variabel. Selv om den besvarer problemstillingen, må tema studeres videre for å kunne gi et mer solid svar.

6.2 Videre studier

Ved videre studier kan det være interessant å se videre på ulikhet i barnehageeffekt mellom landene. Selv om effekten ikke var signifikant ved vår hypotesetest, ville vi kunne forkastet nullhypotesen med 90% sikkerhet. Dersom det i realiteten er ulik effekt av barnehagedeltakelse, kan man se på hvilke faktorer som gjør at effekten eventuelt er større i Norge.

Vi har også merket oss aldersvariabelen, hvor alder gir mindre utslag på leseferdigheter når man kontrollerer for land. Det kan være interessant å undersøke om de eldste elevene som deltok i PIRLS 2001 har lavere effekt av barnehagedeltakelse enn de yngste. I så fall ville den positive barnehageeffekten vært avtakende i alder, som kan indikere at den ikke ville hatt like stor betydning senere i skoleløpet.

Appendiks

	(1) read	(2) read	(3) read	(4) read	(5) read	(6) read
<i>kinderg_att</i>	21.95*** (2.545)	23.56*** (2.639)	15.26*** (2.511)	10.88*** (2.441)	6.517 (3.482)	23.17*** (3.994)
<i>clsize</i>	3.237*** (0.393)		2.685*** (0.376)	1.550*** (0.368)	1.567*** (0.368)	2.698*** (0.376)
<i>clsize_sq</i>	-0.0388*** (0.00593)		-0.0307*** (0.00564)	0.0199*** (0.00549)	-0.0201*** (0.00549)	-0.0309*** (0.00564)
<i>teacher_fem</i>	3.983* (1.887)		0.875 (1.828)	3.003 (1.774)	2.927 (1.774)	0.845 (1.828)
<i>zage</i>	23.45*** (0.739)		21.09*** (0.739)	4.583*** (1.016)	4.593*** (1.016)	21.12*** (0.739)
<i>girl</i>	22.62*** (1.408)		14.67*** (1.383)	13.98*** (1.340)	13.99*** (1.340)	14.67*** (1.382)
<i>par_not_born</i>		-20.02*** (2.629)	-22.28*** (2.553)	-23.60*** (2.474)	-23.83*** (2.447)	-22.05*** (2.553)
<i>early_ability_2</i>		21.30*** (2.617)	15.67*** (2.503)	14.16*** (2.426)	14.21*** (2.426)	15.71*** (2.502)
<i>early_ability_3</i>		44.49*** (2.550)	36.42*** (2.445)	35.33*** (2.380)	35.35*** (2.379)	36.43*** (2.455)
<i>early_ability_4</i>		80.52*** (2.662)	66.92*** (2.589)	65.51*** (2.509)	65.61*** (2.510)	66.94*** (2.588)
<i>pct_disadv26_50</i>		26.08*** (4.106)	27.59*** (4.072)	30.26*** (3.947)	30.43*** (3.948)	27.38*** (4.071)
<i>pct_disadv11_25</i>		15.01*** (3.711)	22.07*** (3.662)	25.79*** (3.552)	25.86*** (3.552)	21.94*** (3.661)
<i>pct_disadv0_10</i>		2.781 (3.585)	20.30*** (3.574)	31.01*** (3.494)	31.14*** (3.495)	20.16*** (3.573)
<i>books100plus</i>		28.30*** (1.564)	27.29*** (1.494)	26.75*** (1.448)	26.76*** (1.448)	39.14*** (4.893)
<i>norway</i>				-53.06*** (2.318)	-60.89*** (4.980)	
<i>kinderg_nor</i>					8.559 (4.875)	
<i>books100_kinderg</i>						-12.99* (5.105)
<i>cons</i>	457.9*** (6.548)	457.4*** (4.839)	407.7*** (7.401)	442.1*** (7.327)	445.9*** (7.635)	400.5*** (7.926)
<i>N</i>	9105	8419	8063	8063	8063	8063
<i>R2</i>	0.166	0.207	0.309	0.351	0.351	0.309

Vedlegg 1: Fullstendig regresjon.

	<i>Koeffisient</i>	<i>Koeffisient r.</i>	<i>Endring</i>	<i>T-verdi</i>	<i>T-verdi r.</i>	<i>Endring</i>
<i>kinderg_att</i>	6.52	6.52	0	1.87	1.89	0.02
<i>cl_size</i>	1.57	1.57	0	4.26	3.89	0.37
<i>clsize_sq</i>	-0.02	-0.02	0	-3.66	-3.33	0.33
<i>teacher_fem</i>	2.93	2.93	0	1.65	1.72	0.07
<i>zage</i>	4.59	4.59	0	4.52	4.25	0.27
<i>girl</i>	13.99	13.99	0	10.44	10.50	0.06
<i>par_not_born</i>	23.83	23.83	0	-9.62	-8.94	0.68
<i>early_ability_2</i>	14.21	14.21	0	5.86	5.11	0.75
<i>early_ability_3</i>	35.35	35.35	0	14.86	13.19	1.67
<i>early_ability_4</i>	65.61	65.61	0	26.14	23.78	2.36
<i>pct_disadv26_50</i>	30.43	30.43	0	7.71	8.20	0.49
<i>pct_disadv11_25</i>	25.86	25.86	0	7.28	7.55	0.27
<i>pct_disadv0_10</i>	31.14	31.14	0	8.91	9.37	0.46
<i>books100plus</i>	26.76	26.76	0	18.48	18.32	0.16
<i>norway</i>	-60.80	-60.80	0	-12.21	-11.07	1.14
<i>kinderg_nor</i>	8.56	8.56	0	1.76	1.58	0.18

Vedlegg 2: Koeffisient og t-verdi med og uten robust verdi. Modell 5.

	read	kinderg~t	clsize	clsize~q	teache~m	zage	girl	par_no~n	early~2	early~3	early~4	pct_d~50
read	1.0000											
kinderg_att	0.1185	1.0000										
clsize	0.1725	0.0395	1.0000									
clsize_sq	0.1453	0.0280	0.9762	1.0000								
teacher_fem	-0.0013	-0.0203	0.0730	0.0871	1.0000							
zage	0.3474	0.1014	0.2516	0.2200	-0.0675	1.0000						
girl	0.1539	-0.0204	0.0033	0.0013	-0.0116	-0.0106	1.0000					
par_not_born	-0.1076	-0.0540	0.0418	0.0407	-0.0151	0.0724	-0.0037	1.0000				
early_abil~2	-0.1950	-0.0028	-0.0153	-0.0129	0.0032	-0.0442	-0.0991	-0.0168	1.0000			
early_abil~3	0.0119	-0.0102	-0.0005	-0.0034	-0.0040	-0.0117	0.0486	0.0223	-0.4737	1.0000		
early_abil~4	0.3241	0.0404	0.0413	0.0381	0.0101	0.1081	0.1344	-0.0035	-0.3720	-0.4375	1.0000	
pct_disadv~50	0.0731	0.0308	0.0305	0.0083	0.0098	0.1202	0.0344	0.0657	-0.0240	0.0018	0.0245	1.0000
pct_disadv~25	0.0771	0.0426	0.0215	0.0093	0.0107	0.1710	-0.0196	-0.0115	-0.0146	0.0198	0.0095	-0.1886
pct_disadv~10	-0.0893	-0.0522	-0.0410	-0.0127	0.0239	-0.2669	-0.0001	-0.1257	0.0298	-0.0207	-0.0281	-0.3858
books_100p~s	0.2197	0.0508	0.0229	0.0092	0.0198	-0.0084	0.0003	-0.1949	-0.0386	-0.0343	0.0804	-0.0392
idcntry_nor	-0.4111	-0.1326	-0.3235	-0.2815	0.0775	-0.7617	-0.0088	-0.0906	0.0306	0.0067	-0.0929	-0.1559
kinderg_nor	-0.3526	0.1834	-0.2934	-0.2554	0.0738	-0.6890	-0.0166	-0.0841	0.0268	0.0093	-0.0834	-0.1423
books100_k~g	0.2294	0.3890	0.0388	0.0218	0.0078	0.0347	-0.0063	-0.1787	-0.0341	-0.0346	0.0861	-0.0261
	pct_d~25	pct_d~10	books_~s	idcntr~r	kinderg~r	books1~g						
pct_disadv~25	1.0000											
pct_disadv~10	-0.7410	1.0000										
books_100p~s	0.0132	0.0428	1.0000									
idcntry_nor	-0.2182	0.3469	-0.0017	1.0000								
kinderg_nor	-0.1876	0.3056	0.0079	0.9029	1.0000							
books100_k~g	0.0280	0.0162	0.8977	-0.0529	0.0700	1.0000						

Vedlegg 3: Korrelasjonsmatrise.

