

## Forord

Vi vil takke Hans Bonesrønning for eminent veiledning og hyggelige samtaler. Takk til Marie H. Sæther for teknisk hjelp i STATA. En takk til Patrik Frideberg Bakken for Latex-support og Torjus Rolid Hagen for behjelpelig korrektur.

Trondheim, 29.05.2015.

*Espen Håkonsløkken Aardalen og Stian Hervik Frantzen*

# Innhold

<b>1</b>	<b>Innledning</b>	<b>1</b>
<b>2</b>	<b>Eksisterende litteratur</b>	<b>5</b>
2.1	Teori . . . . .	5
2.2	Empiri . . . . .	8
2.3	Teorier om vår hypotese . . . . .	11
2.4	Institusjonelle detaljer . . . . .	12
2.4.1	Norske skoleforhold . . . . .	12
2.4.2	Nasjonale prøver . . . . .	13
<b>3</b>	<b>Data</b>	<b>14</b>
3.1	Datasettet . . . . .	14
3.2	Variablene . . . . .	14
3.2.1	Avhengig variabel . . . . .	14
3.2.2	Interessevariabelen . . . . .	17
3.2.3	Kontrollvariable . . . . .	18
<b>4</b>	<b>Økonometrisk tilrettelegging</b>	<b>21</b>
4.1	Kvantitativ metode . . . . .	21
4.2	Regresjon . . . . .	22
4.2.1	Første regresjon . . . . .	22
4.2.2	Andre regresjon . . . . .	24
<b>5</b>	<b>Resultater</b>	<b>27</b>
5.1	Tverrsnittanalyse . . . . .	27
5.2	Første regresjonsmodell . . . . .	27
5.2.1	T-test . . . . .	30
5.2.2	Oppdeling av paritetsvariabelen . . . . .	31
5.2.3	Skoleeffekter . . . . .	32
5.3	Andre regresjonsmodell . . . . .	37
5.3.1	Korrelasjonsanalyse . . . . .	42
5.4	Veien videre . . . . .	43
<b>6</b>	<b>Avslutning</b>	<b>44</b>

# Tabeller

3.1	Standardisert resultat . . . . .	15
3.2	Standardisert resultat for hver søskenparitet . . . . .	17
3.3	Data for kontrollvariabler . . . . .	19
3.4	Utdanningstabell . . . . .	19
3.5	Antall søsken . . . . .	20
5.1	Interesse- og kontrollvariabler . . . . .	28
5.2	Interesse- og kontrollvariabler . . . . .	32
5.3	Interesse- og kontrollvariabler m/skoleeffekt . . . . .	33
5.4	Variabler inkl. skoleeffekter . . . . .	35
5.5	T-test for familiestørrelsen . . . . .	36
5.6	Skoleeffekt . . . . .	37
5.7	Skolefaste effekter . . . . .	38
5.8	Differanse skoleeffekter . . . . .	39
5.9	Forskjell skoleeffekt1 . . . . .	41
5.10	Differanse skoleeffekter u/uteliggere . . . . .	41
5.11	Korrelasjonsmatrise . . . . .	42
5.12	Skoleeffekt1 . . . . .	43

# Figurer

3.1	Tetthetsfunksjon for elevprestasjon . . . . .	16
3.2	Avhengig- og interessevariabel . . . . .	18
4.1	Skoleeffekt . . . . .	25
5.1	Tetthetsfunksjon for skoleeffekten . . . . .	37
5.2	Tetthetsfunksjon for skoleeffektene . . . . .	39
5.3	Observasjoner skoleeffekt1 . . . . .	40

# Kapittel 1

## Innledning

Effekter av søskenparitet, plassering i søskenflokk, er noe som har blitt diskutert i lang tid. Fra gammelt av har det vært tradisjon for at plassering i søskenflokk gir deg enkelte fordeler og/eller ulemper. Dette gjelder spesielt hvis du er født først, gjerne hvis du er førstefødte sønn. Førstefødte var familiens arving og den som skulle bære slekten, bedriften eller kongeriket videre. Var du født senere; andre-, tredje- eller fjerdefødt, arvet du lite eller ingen ting med mindre dine eldre søsken gikk bort. Besitter førstefødte egenskaper som gjør han/hun mer egnet for ansvar og beslutningstaking, eller kan en andre-, og tredjefødte gjøre en like god jobb? Helt tilbake til Bibelen og norrøn mytologi kan vi finne tekster som viser at folk lenge har hatt ideer om hvordan plasseringen i søskenflokk påvirker egenskaper og personlighetstrekk hos folk.

I moderne tider har vi i større grad kvittet oss med synsingen og filosoferingen, og begynt å studere reelle data for å påvise eller tilbakevise søskenrekkefølgeeffekt. En av de første publiserte forskningsartiklene på søskenrekkefølgeeffekten sto Galton (1874) bak da han observerte medlemmer av Royal Society i London. Han fant at det var klart flest førstefødte og enebarn blant medlemmene. Clarke (1916) utførte en lignende studie på amerikanske menn med høy utdanning med lignende resultat.

Søskenrekkefølgeeffekten på intelligens har blitt diskutert mellom forskere lenge. Schooler (1972) mente å kunne tilbakebevise effekten i en studie av amerikanske menn på sekstitallet.

Studien fikk navnet *Birth order effects: Not here, not now!*. Konklusjonen var tydelig der Schooler ikke observerte noen søskenrekkefølgeeffekt. Året etter kom Belmont & Marolla (1973) med en studie av nederlandske 19-åringer og fant evidens for at det finnes en negativ sammenheng mellom søskenrekkefølge og intelligens. I senere tid har søskenrekkefølgeeffekten blitt mer og mer anerkjent, spørsmålet nå er ikke om effekten er der eller ikke, men hva som forårsaker effekten og hva effekten har å si for et individs produktfunksjon. Psykologer har vært mest opptatt av årsak, mens økonomer har fokusert mer på resultater som medfølger effekten. Produktfunksjonen viser hvilke faktorer som påvirker et individs produksjon. Personlige egenskaper spiller stor rolle for et individs produksjon, og søskenrekkefølgeeffekten påvirker de personlige egenskapene til en person (Black, Devereux & Salvanes 2005). Derimot vil også en effekt av skolen være naturlig å ta med i produktfunksjonen (Rivkin, Hanushek & Kain 2005)

I skrivende stund er det fortsatt uenighet om søskenrekkefølgeeffekten. Hanushek (1992) oppdaget i en studie av elevprestasjoner at når familiestørrelsen ble holdt konstant, var det ingen signifikante søskenrekkefølgeeffekter. Andre artikler har anerkjent effekten og fokuserer mer på veien videre og hvilke årsaker som ligger til grunn for effekten. Det blir benyttet ulike mål, eller avhengige variabler, når effekten måles, der vi blant annet har IQ (Bjerkedal, Kristensen, Skjeret & Brevik 2007), antall års utdanning (Black et al. 2005) og elevprestasjon (Bonesrønning & Massih 2011).

I vår oppgave lager vi en produktfunksjon for elever som gikk i 5. klasse i 2007, for å undersøke søskenrekkefølge, elevprestasjon og skolekvalitet. Vi har resultatet på matematikkdelen fra nasjonale prøver det året som avhengig variabel. Vi bruker familiære og individuelle trekk som forklaringsvariabler, med søskenparitet som interessevariabelen.

Hotz & Pantano (2015) danner seg en hypotese som sier at en av mekanismene som skaper søskenrekkefølgeeffekter er strategisk oppdragelse og svarene barna gir på oppdragelsen. De tar utgangspunkt i litteraturen som argumenterer for at senerefødte søsken er mindre ansvarsbevisste og mindre selvstendige enn sine eldre søsken. Derfor vil førstefødte generelt prestere bra på skolen. De yngre søsknene observerer at førstefødte får straff for å stjele, slå, lyve osv., men observerer sjeldent at foreldrene straffer dårlige skoleresultater. Derfor kan

de yngre søsknene tillate seg relativt dårligere elevprestasjoner, fordi de forventer liten eller mild straff fra foreldrene. I denne oppgaven blir modellen til Hotz og Puntano brukt med en utvidelse. Gitt at strafferegimet til Hotz og Puntano gir opphav til søskenrekkefølgeeffekter, der de yngre søsknene kan tillate seg dårligere elevprestasjoner, ønsker vi å undersøke om skolekvaliteten har mer å si for senerefødte søsken versus førstefødte søsken.

I boken *The Birth Order Book* sier Leman (2009) at sistefødte vil oppleve litt mer avslappede foreldre enn den førstefødte. Skuldrene er senket fordi det gikk bra den/de første gangen/gangene. Dette gjenspeiles i barna, de er ofte frisinnede og bekymringsfri og trenger mindre sikkerhet i livet i forhold til de strukturerte og ansvarlige eldre søsknene. Dette kan være en grunn til at skolekvaliteten er viktigere for senerefødte. Ved høy søskenparitet øker sannsynligheten for at du er sistefødt og du har derfor et "last-child"-handicap definert av Zajonc & Markus (1975). Siden du ikke har yngre søsken mister du fordelene ved å være lærer og "sjef" for dine yngre søsken. Dette kan føre til at sistefødte blir mer selvsentrert, mindre ansvarsfull og tenker ikke konsekvenser på samme måte som førstefødte.

Ved å kombinere det som er beskrevet ovenfor kommer vi frem til at barn med høy søskenparitet kan ha foreldre som gir friere tøyler, noe som kan føre til at barnet har et mer avslappet forhold til foreldre og lærere. Barnet tillater seg selv dårligere elevprestasjoner fordi forventningen om straff er lav, og jo senere du er født jo større er sannsynligheten for at du har et "last-child"-handicap. Alle disse årsakene motiverer hypotesen vår om at skolekvaliteten spiller større rolle for senerefødte. Dette vil si at vi har grunnlag for å tro at variasjonen i elevprestasjonene til en førstefødt vil variere lite på tvers av skolene. Selv om skolen har lav kvalitet vil fortsatt en førstefødt prestere godt på grunn av de kvalitetene en førstefødt har fått igjennom oppveksten. Vi kan forvente oss å se stigende variasjon på elevprestasjonene jo lengre ut i søskenflokkene vi beveger oss. En som er andre- eller tredjefødt kan i utgangspunktet være like flink, eller flinkere, på skolen enn førstefødte, men på grunn av forholdene beskrevet over vil en lærer som både motiverer til gode prestasjoner og straffer dårlige prestasjoner være viktigere. Vi antar at ikke alle skoler har like gode lærere, og at lærertettheten varierer på tvers av skolene. I oppgaven vår ønsker vi å undersøke om elevprestasjonene til elevene med høy søskenparitet varierer mer med skolekvaliteten i forhold til elevene med

lavere søskenparitet.

## Hypotese

Debatten rundt hvorvidt søskenrekkefølgeeffekten er reell eller ikke diskuteres fortsatt, så det er nødvendig for oss å undersøke om vi har evidens for søskenrekkefølgeeffekt i datasettet vi bruker. Vi må derfor undersøke en hypotese som sier at det er negative effekter forbundet med søskenparitet blant elevene i datasettet vårt. Deretter vil vi undersøke hovedhypotesen som sier at skolekvalitet er viktigere for elever med høy søskenparitet. Dette vil vi undersøke ved å se på variasjonen til skoleeffektene for de ulike søskenparitetene. Til slutt undersøker vi korrelasjonen mellom skoleeffektene for de ulike søskenparitetene. Det for å se om en skole som påvirker førstefødte positivt også vil påvirke andrefødte og/eller tredjefødte positivt.

Oppgaven blir et bidrag i grunnforskningen på søskenrekkefølgeeffekt blant unge barn og vil være en del av puslespillet som gir oss bedre forståelse av individers produktfunksjon. Ved å inkludere skole som disiplininstans i modellen til Hotz & Pantano (2015) blir oppgaven et bidrag til å forstå hva som påvirker elevens produktfunksjoner. Oppgaven stiller seg i rekken av tidligere forskning som har funnet evidens for søskenrekkefølgeeffekt og går et steg videre ved å se på hvordan skolekvalitet påvirker elevprestasjoner for elever med ulik paritet.

## Oppgavens oppbygging

Oppgaven fortsetter med Kapittel 2 som er en gjennomgang av relevant litteratur om søskenrekkefølgeeffekt og elevprestasjoner hvor nasjonale prøver som sammenligningsgrunnlag også blir diskutert. Kapittel 3 tar for seg dataene vi har fra nasjonale prøver for 5. klassinger i 2007. Kapittel 4 presenterer den økonometriske tilretteleggingen, hvordan vi angriper hypotesene vi vil undersøke og mulige problemer som kan dukke opp. Her deles hypotesene opp i to der vi først undersøker om det er søskenrekkefølgeeffekt, deretter undersøker vi skoleeffektene. Videre vil resultatene bli presentert i kapittel 5 og i likhet med kapittel 4 deles resultatene opp i to. Til slutt, i kapittel 6, konkluderer vi med funnene vi gjør i undersøkelsen av hovedhypotesen vår.



# Kapittel 2

## Eksisterende litteratur

### 2.1 Teori

Søskenrekkefølgeeffekten er en effekt det har blitt forsket mye på fra forskjellige studieretninger. Psykologiske teorier fokuserer på hvordan søsken utvikler forskjellige personligheter og knytter dette opp mot søskenrekkefølgeeffekten. En interessant teori om søskenrekkefølgeeffekten ble lagt fram av Belmont og Marolla (1963) hvor de kommenterer det intellektuelle miljøet i en familie. Det argumenteres for at søskenrekkefølgeeffekten avhenger av familiestørrelse, også kalt familiestørrelseeffekten. Her er det det intellektuelle miljøet innad i en familie som brukes som argument for søskenrekkefølgeeffekten. Zajonc & Markus (1975) har videreutviklet denne teorien og den har senere blitt kalt konfluensteorien. Konfluensteorien tar utgangspunkt i en familie hvor det eksisterer et gitt intellektuelt nivå. Vi antar at begge foreldrene har intelligens på 100 (IQ). Når førstefødte kommer inn i familien, med gitt intelligensnivå 0, vil gjennomsnittsintelligensen i familien bli;  $100 + 100 + 0 / 3 = 67$ . Etter hvert som årene går og førstefødte har blitt noen år og oppnådd intellektuelt nivå lik 40 blir andrefødte født. Det intellektuelle nivået i familien vil nå bli dratt ned;  $100 + 100 + 40 + 0 / 4 = 60$ . Hele det intellektuelle miljøet har blitt forverret og dette vil påvirke den intellektuelle utviklingen til begge barna. Hvis vi nå fortsetter noen år fremover i tid og antar at et tredje barn blir født i familien når førstefødte har intellektuelt nivå på 50 og andrefødte 30, vil intelligensnivået i

familien bli seende slik ut:  $100 + 100 + 50 + 30 + 0/5 = 56$ . Slik fortsetter det hvis vi legger til flere nyfødte. Ifølge denne teorien vil store familier bli sett på som miljøer med lavere intellektuelle nivåer fordi større del av familien bidrar med lav absolutt intelligens.

Det er viktig å ta hensyn til effekten av stor familie når vi undersøker søskenrekkefølgeeffekten, for om du har høy søskenparitet er også familien stor. Videre argumenterer Zajonc & Markus (1975) for at aldersforskjellen mellom søsken er viktig i konfluensteorien. Blir du født inn i en familie hvor det er stor aldersavstand opp til dine søsken er det intellektuelle nivået i familien høyere enn om du er født tidsmessig tettere til dine søsken. Eldre søsken utvikler sine intellektuelle kapasiteter ved å fungere som en mentor for sine yngre søsken (Leman 2009). Dette er en positiv effekt alle barna i en søskenflokk vil nyte godt av, bortsett fra sistefødte.

Vi finner også teorier med biologisk tilnærming. Det er kanskje ikke bare oppdragelse og miljø som bestemmer personlighetstrekkene til barn, gener kan også spille en viktig rolle. Clarke (1916) prøvde seg på en hypotese om at livmoren favoriserer førstefødte og at effekten er avtagende for resterende søsken. Behrman (1988) påpeker at barn med høyere søskenparitet stort sett har eldre mødre. Eldre mødre føder generelt barn med relativt lavere fødselsvekt, og dette favoriserer førstefødte. Men på en annen side favoriseres yngre søsken fordi foreldrene har erfaring med barneoppdragelse når de blir født. De biologiske argumentene er sprikende og motsigende.

En økonomisk teori har prøvd seg på å forklare søskenrekkefølgeeffekten, nemlig ressursutjevningsteorien. Her antar vi at foreldrene har et gitt nivå med ressurser og barna konkurrerer seg i mellom om ressursene. Færre søsken betyr færre konkurrenter (Blake 1981). Aldersavstanden kan også være viktig her (Powell & Steelman 1990). Barn med lav søskenparitet har relativt færre konkurrenter, mens sistefødte tjener på familiens antatte inntektsvekst over tid (Downey 2001). Price (2008) gjorde en undersøkelse der han så på antall timer ”kvalitetstid” med foreldrene hvert barn får. Studien viser at generelt sett får førstefødte mer kvalitetstid både med mor og far relativt til andrefødte. For å oppsummere nyter førstefødte av mer tilgjengelige ”foreldreressurser” i form av både tid og penger.

Økonomisk forskning har vært interessert i hvilke faktorer som påvirker et barns produkti-

vitet. Ønsket har vært å kunne lage en produktfunksjon med forklaringsvariabler som kan determinere ”produksjonen” til en person. Black et al. (2005) argumenterer for at familie og oppvekst har stor påvirkning på produktfunksjonen til en person. Familiestørrelse er en av disse faktorene man tror kan ha stor påvirkning på produksjon. Becker (1960) introduserte oss for kvantitet-kvalitetmodellen hvor hypotesen gikk ut på å forklare den observerte negative korrelasjonen mellom familieinntekt og familiestørrelse. Her belyses en interaksjon mellom kvantitet og kvalitet i budsjettbetingelsen som fører til økende marginalkostnader på kvalitet med hensyn på familiestørrelse. Dette genererer et trade-off mellom kvantitet og kvalitet. Empirien tilsier at denne effekten eksisterer og at barn fra store familier generelt har lavere utdanningsnivå, men hva som er årsakene bak denne effekten er ikke avklart. For er det slik at ”kvaliteten” på barnet minker når søskenflokket utvides eller er denne effekten sterkere knyttet opp mot foreldrene? Kan det være slik at store familier er store fordi det er et valg foreldrene har tatt og sier derfor mer om foreldrene enn barna? Hvis alle ble pålagt å få samme antall barn, ville vi da sett søskenrekkefølgeeffekt eller ville den blitt ”spist opp”?

I denne oppgaven vil vi ta utgangspunkt i modellen til Hotz & Pantano (2015) i artikkelen *Strategic Parenting, Birth Order and School Performance*, som foreslår at søskenrekkefølgeeffekten på elevprestasjon kommer av det de kaller strategisk oppdragelse. Artikkelen hevder at foreldre er strengere med de eldre barna, deretter løsner de grepet gradvis på de yngre. Det førstefødte barnet vil muligens oppleve restriksjoner på bruk av TV og datamaskin og foreldrene vil være nøyere på oppfølging av lekser. Når foreldrene får flere barn vil de observere at det faktisk gikk bra med førstefødte og er derfor mer avslappet og ikke fullt så strenge med de yngre søsknene. Ved å være strenge med det eldste barnet setter også foreldrene et eksempel for de yngre.

Modellen til Hotz og Puntao, som først ble utviklet av Hao, Hotz & Jin (2008), er et ”endelig” spill hvor deltakerne er foreldrene og alle barna i søskenflokket. I alle rundene av spillet,  $t, t = 1, 2, \dots, N$ , observerer foreldrene og de yngre søsknene spillhistorikken mellom foreldrene og det eldste barnet. De yngre søsknene observerer sine eldre søskens adferd og hvilken straff/belønning denne adferden medfører. I vårt tilfelle er det reaksjonen fra foreldrene og de yngre søsknene, hvis de eldste barna i familien presterer dårlig på skolen, som er interessant.

I den første runden av spillet vet ikke noen av barna hva foreldrenes reaksjon på dårlige skoleresultater vil være. Foreldrene kan enten velge den tøffe formen for oppdragelse, hvor straff for dårlige skoleresultater er på sin plass, eller den mer avslappede typen som sverger til oppmuntring fremfor straff. Vi definerer  $u_1$  som de yngre barnas tro, eller sannsynlighet, på at foreldrene er av den strenge typen og  $1 - u_1$  som sannsynlighet for at foreldringer er av den ikke-strenge typen. Yngre søsken som observerer eldre søsken som presterer godt i skolen, og derfor aldri får noen straff, kan anta at det ikke er noen straff for dårlige skoleresultater.  $u_1$  vil da gå ned til null i denne runden og alle fremtidige runder av spillet.

Hao et al. (2008) påpeker at førstefødte aldri har fått observere foreldrenes strafferegime og er usikker på hva slags type, strenge eller ikke-strenge, de er. Derfor vil de sikre seg mot straff og føle mer press på å prestere godt på skolen. Videre argumenteres det for at i denne fasen av spillet vil det eldste barnet bli straffet for dårlige skoleresultater uavhengig av foreldretypen. Strenge foreldre vil straffe av prinsipp og ikke-strenge vil straffe for å sette et eksempel for resten av søskenflokket. Hotz og Puntano beskriver at effekten av å straffe eldre søsken for å sette eksempel for de yngre er en avtagende effekt. Det kommer til slutt til et punkt der straff for eksempelets skyld har lavere effekt enn hva det koster for foreldrene å ta fra barna frihet og tid. Ved høy søskenparitet er derfor sannsynligheten mindre for at du har erfart at eldre søsken blir straffet for dårlige skoleresultater. Siden straff ikke er observert vil det være nærliggende for yngre søsken å teste hvor langt strikken kan tøyes før det blir straff. De yngre søsknene tillater seg derfor dårlige skoleresultater fordi forventningen om straff er lav og usikker. Dette kan være en av mekanismene i søskenrekkefølgeeffekten.

## 2.2 Empiri

En rekke nyere empiriske studier viser at søskenrekkefølgeeffekten eksisterer. Disse studiene har brukt litt forskjellige utfallsvariabler. Black et al. (2005) gjorde en studie av norske familier for å undersøke familiestørrelseeffekten og søskenrekkefølgeeffekten på antall år med utdanning som utfallsvariabel. Studien finner en sterk negativ familiestørrelseeffekt på antall år utdanning, men når søskenrekkefølge blir inkludert reduseres familiestørrelseeffekten.

Black et al. trodde at trenden i industriland med mindre familier skulle slå positivt ut på barns prestasjoner, og selv om gjennomsnittsbarnet presterer bedre, finner studien kun små effekter av familiestørrelse på førstefødte. På den andre siden finner studien sterke søskenrekkefølgeeffekter på antall år utdanning. Studien utfyller "antall år utdanning"-variabelen med inntekt, fast eller deltidsstilling, og om personene i undersøkelsen fikk barn som tenåring, som tilleggsvariabler. Videre finner de at søskenrekkefølgeeffekten spiller en større rolle spesielt for kvinner. Er du kvinne og født lengre ut i søskenflokket tjener du mindre, har mindre sjanse for fulltidsjobb og større sjanse for å bli forelder allerede som tenåring. Menn som er født lengre ut i søskenflokket har generelt lavere lønn, men har like stor sannsynlighet som sine eldre søsken til å jobbe fulltid. Resultatene viser også at søskenrekkefølgeeffekten er høy også hos familier med høyt utdannet mor. Dette sammen med de svake familieeffektene tilsier at de finansielle restriksjonene kanskje ikke er så signifikante som Black et al. tidligere antok. Et viktig poeng som kommer frem i studien er at det tidligere har vært for mye fokus på de mellomfamiliære forskjellene, men at det også er viktig å ta med forskjellene som varierer innad i hver familie. Black et al. avslutter studien med å si at hvis funnene i studien stemmer må vi ta en ny vurdering på hvilke variabler vi vil ha med i produktfunksjonen til barn. De mener de varierende resultatene forskjellige studier får tilsier at feltet trenger mer arbeid.

Blant økonomer er det produktfunksjonen og estimeringen av den som er interessant. Problemet med estimering av søskenrekkefølgeeffekten er at det ofte gir skjevheter. En milepæl innen forskningen ble nådd da Bjerkedal et al. (2007) undersøkte effektene av søskenrekkefølge både innen og mellom familier i et nærmest komplett utvalg av norske 18-19-åring. Her ble IQ brukt som utfallsvariabel. Tidligere hadde man studert variasjonene mellom familiene, men nå ble de familiefaste effektene, "within"-variasjonen, undersøkt. Ved å gjøre dette klarte Bjerkedal et al. å kontrollere for mellomfamilievariabler som mors alder, mors alder i forhold til barna, fars alder, fars inntekt, regionalt fødested og fødselsvekt. Studien konkluderer med evidens for søskenrekkefølgeeffekt på IQ, men oppfordrer til videre testing av søskenrekkefølgeeffekt i andre datasett.

I en annen studie av Kristensen & Bjerkedal (2007) ble igjen sammenhengen mellom søs-

kenrekkefølge og intelligens undersøkt. Her konkluderes det med at sammenhengen mellom intelligens og søskenrekkefølge avhenger av den sosiale rangen innad i familien. Generelt er det førstefødte som er "sjefen" i søskenflokket, mens de påfølgende søsknene deretter følger på rangstigen. I noen tilfeller, som ved dødsfall, kan yngre søsken tre frem i lederrollen og vil derfor agere som en førstefødt.

Bonesrønning & Massih (2011) argumenterer for at tid med mor og far påvirker elevprestasjonen (samme utfallsvariabel vi bruker). Det viser seg at førstefødte er negativt påvirket av avstanden til andrefødte. Kortere avstand er bra for førstefødte, mens andrefødte derimot er upåvirket av avstanden opp til førstefødte. Dette på grunn av foreldrepermisjonen vil førstefødte få lengre sammenhengende periode med foreldrene hjemme når andrefødte er født nærmere førstefødte. McLanahan (2004) påpeker at det er større fordel for førstefødte i familier med høyt utdannet mor da høyt utdannede mødre kan gi barnet flere ressurser i form av tid og penger relativt til mødre med lavere utdanning. Dette favoriserer barn med høyt utdannede mødre. Denne effekten antas å ha sammenheng med mors alder ved fødsel, siden høyt utdannede kvinner generelt venter lenger med å starte familie. Kantarevic & Mechoulan (2006) viser i sin studie viktigheten av å kontrollere for mors alder ved fødsel. De argumenterer for at lav alder på mor ved fødsel kan ta bort noen av fordelene ved å være førstefødt. I tilfeller der moren er ung, er fødselen ofte dårlig planlagt, eller ikke planlagt i det hele tatt, og det kan oppstå utfordringer både sosialt, økonomisk og familiært. Men selv når studien til Bonesrønning og Massih kontrollerer for mors alder ved fødsel, finner den klare fordeler ved å være førstefødt i en familie med høyt utdannet mor.

Ikke alle studier konkluderer med evidens for søskenrekkefølgeeffekt. Hanushek (1992) konkluderer i en studie av afro-amerikanske lavinntektsfamilier i USA, at når familiestørrelsen holdes konstant finnes ingen signifikante søskenrekkefølgeeffekter for familier med to, tre eller fire barn. Studien bruker elevprestasjon som utfallsvariabel og argumenterer for at den eneste fordel en elev har når han eller hun er født først, er at det er større sannsynlighet for at han eller hun kommer fra en liten familie. Studien konkluderer med signifikante negative effekter forbundet med stor familie.

Stinebrickner & Stinebrickner (2008) viste i en studie at for yngre elever har innsats på

skolen veldig mye å si for elevprestasjonen, mens selve intelligensen har mer å si når elevene kommer høyere opp i skolesystemet. Dette underbygger modellen til Hotz og Puntano som tar utgangspunkt i at eldre søsken legger ned større innsats fordi de forventer straff ved dårlige resultater, mens de yngre prøver å tøyne strikken og nedlegge lavest mulig innsats for å teste ut strafferegimet til foreldrene.

## 2.3 Teorier om vår hypotese

Som grunnlag for analysen velger vi å utvide en modell utviklet av Hotz & Pantano (2015).

Vi utvider modellen med en produktfunksjon for elevene gitt ved:

$$Resultat = f(\text{skoledisiplinering}, \text{familiedisiplinering}) \quad (2.1)$$

Her er elevprestasjonen, *Resultat*, et produkt av funksjonen, *f*, som bestemmes av disiplineringen eleven får fra hjemmet og skolen. Vi antar her at disiplinering, både fra hjemmet og fra skolen, er positivt, men med avtakende effekt og at skole- og familiedisiplinering fungerer som substitutter. Når familiedisiplineringen minker, må skoledisiplineringen øke tilsvarende for et gitt nivå på elevprestasjonen. Denne økningen i skoledisiplineringen er det usikkerhet rundt med tanke på hvordan skolene responderer på problemet. Hvordan de forskjellige skolene håndterer slike situasjoner kan variere mye fra skole til skole. Vi kan tenke oss at skoledisiplinering kan skrives som en funksjon der:

$$SD = g(\text{laererkompetanse}, \text{ressursinnsats}, \text{osv.}) \quad (2.2)$$

*SD* er skoledisiplineringen som er produkt av funksjonen *g*. Faktorene som påvirker *g* vil variere mellom skolene og derfor vil elever oppleve forskjellige ”disiplinsregimer”. Vi mener at dette kan gi grunnlag for større variasjon i elevprestasjoner hos elever med høyere søskenparitet. Et annet problem ved å analysere skolerresultater på denne måten er om en god skole også er en god skole for elever med forskjellig søskenparitet.

Det kan tenkes at ressursutjevningsteorien påvirker andre- og tredjefødte ved at de ikke får like mye oppmerksomhet og ressurser som førstefødte og er derfor mer avhengig av skolen som en bidragsyter til bedre elevprestasjoner. Price (2008) foreslår at førstefødte får større del av ressursene/oppmerksomheten også når søskenflokken blir eldre. Kanskje er dette svaret på hypotesen vår, at skolekvaliteten er viktigere for senerefødte, fordi skolen opptar større andel av den totale oppmerksomheten individet får i forhold til den førstefødte. Bonesrønning & Massih (2011) argumenterer for at barn med høyere søskenparitet får mindre tid med foreldrene hjemme. Dette kan være med på å gi opphav til søskenrekkefølgeeffekten. Her igjen må vi tenke at ved høyere søskenpariteten er det større sannsynligheten for at du er sistefødt som resulterer i mindre tid med foreldrene relativt til dine eldre søsken. Hvis vi antar at mindre kvalitetstid med foreldrene virker negativt, taler dette for at prestasjonene til en tredjefødt vil variere mer enn for første- og andrefødte.

Konfluensteorien, som sier at større familier har lavere intellektuelt klima i hjemmet, kan brukes som en forklaring på søskenrekkefølgeeffekt. Siden en tredjefødt er født inn i et relativt lavere intellektuelt klima enn førstefødte, vil skolen ha større betydning på elevprestasjonen. Her oppstår et problem som er vanlig i undersøkelser av søskenrekkefølgeeffekt, nemlig om den negative effekten skyldes at eleven har høy søskenparitet, eller at eleven kommer fra en stor familie.

## 2.4 Institusjonelle detaljer

### 2.4.1 Norske skoleforhold

I Norge har vi obligatorisk skole fra 1-10. klasse. Skolene driftes og eies av kommunene som igjen får finansiering fra staten. Privatskoler underviser under 3% av norske elever, noe som tilsier at privatskole ikke er et alternativ til offentlig skole i Norge (Leuven, Oosterbeek & Rønning 2008). Skolene har geografiske nedslagsfelt som gjør at foreldrene i utgangspunktet ikke kan velge hvilken skole barna skal gå på, og derfor går ofte søsken på samme skole. Strøm (2004) påpeker at den sterke integrerings- og rettferdighetspolitikken som føres i



Norge medfører at elever i samme klasse skal bli behandlet likt og bli undervist samlet i klassen. Dette for at alle elevene skal gjennomføre de obligatoriske ti årene på like mange år, i motsetning til andre land hvor ”toppelevne” blir plassert i egne klasser og får muligheten til å fullføre grunnskolen på kortere tid.

### 2.4.2 Nasjonale prøver

Bondevik II-regjeringen innførte våren 2004 obligatoriske nasjonale prøver for grunnskolen som en prøveordning. I starten var ideen at prøvene skulle holdes ved avslutningen av hovedtrinnene i den tiårige grunnskolen; 4. 7. og 10. trinn. Det ble gjennomført nasjonale prøver igjen i 2005, hvor arbeidet rundt prøvene ble evaluert, og i 2006 vedtok stortinget at nasjonale prøver skulle starte opp i ny form i 2007 (Seland, Vibe & Hovdhaugen 2013).

For forskere og andre som skal bruke dataene som produseres av prøvene er det viktig at den faktiske virkeligheten blir gjenspeilet. Elstad (2009) kommenterte at lærere som har elever som presterer dårlig på nasjonale prøver blir uthengt i media og kritisert for dårlige læringsmetoder. Dette kan føre til at lærere i enkelte klasser driller elevene på oppgaver som typisk kommer på de nasjonale prøvene for å bedre klassens resultater. Skoler som driver med ”teaching to the test” kan forskyve andre fag fra læreplanen som vil gi skolene bedre resultater på nasjonale prøver på kort sikt, men på lang sikt har skolens elever mangler med seg i selve kunnskapsgrunnlaget (Seland et al. 2013). Forholdene som er beskrevet her kan gi feil data som igjen kan gi feilaktige resultater når data fra nasjonale prøver blir brukt til forskning.

Denne oppgaven bruker nasjonale prøver som mål på elevprestasjoner og det er derfor nærliggende for oss å stille spørsmål om resultatene som kommer frem her. Det kan tenkes at søskenpariteten til eleven kan gjenspeile hvilken tilnærming eleven har til prøven. Den nasjonale prøven påvirker ikke karakterene til elevene og man kan derfor stille seg spørsmålet om elevene tar den seriøst eller ikke. På bakgrunn av personlighetstrekkene til søskenparitetene, vil kanskje tidligerefødte derfor ta prøven mer seriøst enn senerefødte?

# Kapittel 3

## Data

### 3.1 Datasettet

Datamaterialet som benyttes i oppgaven er et datasett fra nasjonale prøver i grunnskolen i Norge. Dette materialet har vi hentet fra statistisk sentralbyrå. Det inneholder observasjoner fra alle kommuner i Norge og her finnes data for de ulike elevenes resultater i realfagene matte, norsk og engelsk. I tillegg finnes det informasjon om elevenes alder, kjønn, klassetrinn, etnisk bakgrunn, mors/fars inntekt, mors/fars utdanning, antall søsken, søskenparitet, kommune, skole og mye mer.

### 3.2 Variablene

#### 3.2.1 Avhengig variabel

Vår avhengige variabel, *Resultat*, tar for seg resultater fra nasjonale prøver i matematikk for 5. klasseelever i år 2007. Vi har generert et standardisert matteresultat for denne variabelen, dette for å vise et forenklet bilde av resultatene. Det vil si en normalfordeling med forventningsverdi lik 0 og standardavvik lik 1. Vist med ligning  $N(\mu, \sigma) = N(0, 1)$ . Forventningsverdien,  $\mu$ , på 0 vil si at gjennomsnittet eller forventet resultat for eleven er 0.

Standardavviket,  $\sigma$ , beskriver sannsynligheten for at elevens resultat er innenfor et gitt intervall. Det er en sannsynlighet på 68,2% at en gitt observasjon ligger innenfor en avstand på ett standardavvik fra gjennomsnittet. Om observasjonene ligger innenfor en avstand på to standardavvik fra gjennomsnittet er sannsynligheten på 95,4%. Hvis vi med andre ord tar ut en tilfeldig elev fra analysen kan vi si at den personen med 68,2% sannsynlighet har et resultat på den nasjonale prøven mellom -1 og 1.

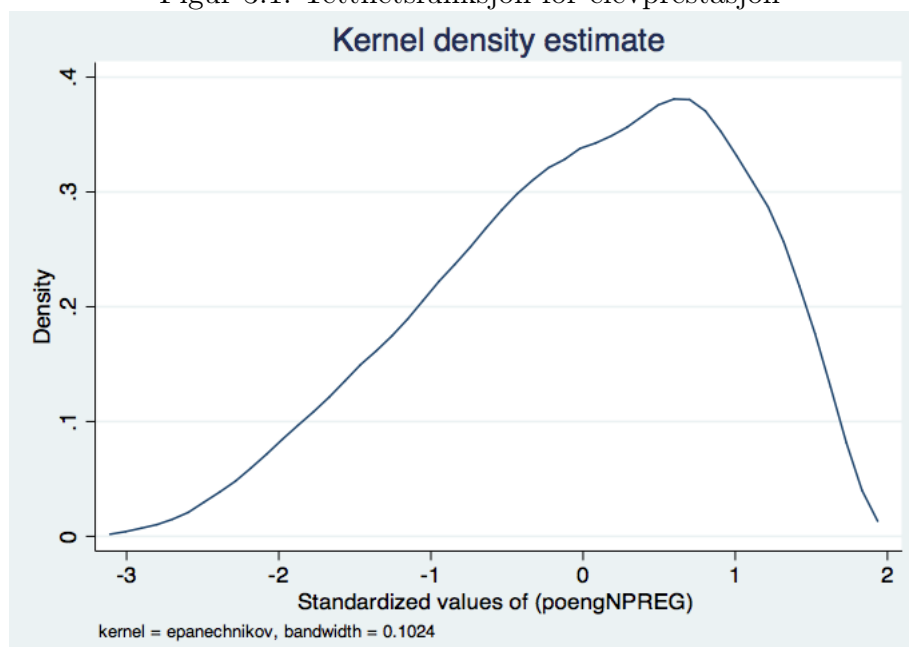
Tabell 3.1: Standardisert resultat

Variabel	Antall obs.	Mean	Std. Dev.	Min	Max
Resultat	52573	0	1	-3.007054	1.83368

Tabell 3.1 viser at vi totalt har 52 573 observasjoner og dermed like mange elever med registrerte resultater på prøven. Matteredultatet til den enkelte eleven varierer fra -3, som da er det dårligste resultatet, til 1,83 som er det høyeste. Vi vil hovedsakelig gjennomføre to analyser. I den første undersøker vi sammenhengen mellom avhengig variabel og søskenparitet, altså hvordan søskenrekkefølgen påvirker resultatet. I den andre analysen undersøker vi hvordan skoleeffekten påvirker resultatene til de ulike søskenparitetene. I begge analysene vil *Resultat* bli brukt som avhengig variabel.

Videre skal vi se hva slags fordeling vi har. Da absoluttverdien til minimum er større enn maksimum,  $|-3| > |1,83|$  mistenker vi skjev fordeling. Viser fordelingen i figur 3.1.

Figur 3.1: Tetthetsfunksjon for elevprestasjon



Grafen viser hvordan resultatene er fordelt på de totalt 52 573 elevene. X-aksen viser resultatene på den nasjonale prøven og Y-aksen viser antall elever på de ulike resultatene. Høy Y-verdi betyr stor andel elever på det gitte resultatet. Vi ser tydelig av figuren at fordelingen ikke er normalfordelt, da de fleste elever ender opp med et resultat på ca. 0,8 standardavvik over gjennomsnittet (høyeste punkt på figuren). Ved en normalfordeling vil gjennomsnittet ligge rett under toppunktet og fordelingen vil være lik på begge sider, noe som ikke er tilfelle ved vår figur hvor gjennomsnittseleven har resultat lik 0. Fordelingen har en "lang venstrehale" fram til toppunktet, 0,8 standardavvik over gjennomsnittet, for deretter å falle brått ned til 1,83 standardavvik over gjennomsnittet. Det forteller noe om hvordan fordelingen av resultatene på elevenes nasjonale prøver er. De som har gjort det godt<sup>1</sup> har mer samlede resultater. Det er mange elever med ca. like gode resultater. De som har gjort det dårlig<sup>2</sup> varierer i større grad, og det er ikke i like stor grad elever som havner på likt resultat som hos de gode elevene. Det er større sprik i resultatene, noe vi ser av ytterpunktene -3 mot 1,83, og den lange venstrehalen illustrerer dette.

<sup>1</sup>Over gjennomsnittet

<sup>2</sup>Under gjennomsnittet

### 3.2.2 Interessevariabelen

Vi ønsker å se om resultatet på den nasjonale prøven avhenger av søskenparitet. Har man også her fordel av å være førstefødt slik det er beskrevet i flere tidligere nevnte forskningsartikler? For å undersøke dette må vi se på variabelen *Paritet* som beskriver elevens nummer i søskenflokk. *Paritet*=1 betyr førstefødt, 2 andrefødt, 3 tredjefødt osv.

Av tabell 3.2 ser vi at det er registrert flest førstefødte med 39,12% av alle observasjonene. Andrefødte følger tett etter med hele 36,7% og mengden av elevene som er tredjefødte er på 18,03%. Sannsynligheten for at en vilkårlig elev enten er første-, andre- eller tredjefødt er dermed på hele 93,84% og vi vil derfor videre i oppgaven kun bruke disse verdiene for vår interessevariabel da de dekker såpass stor grad av observasjonene.

Tabell 3.2: Standardisert resultat for hver søskenparitet

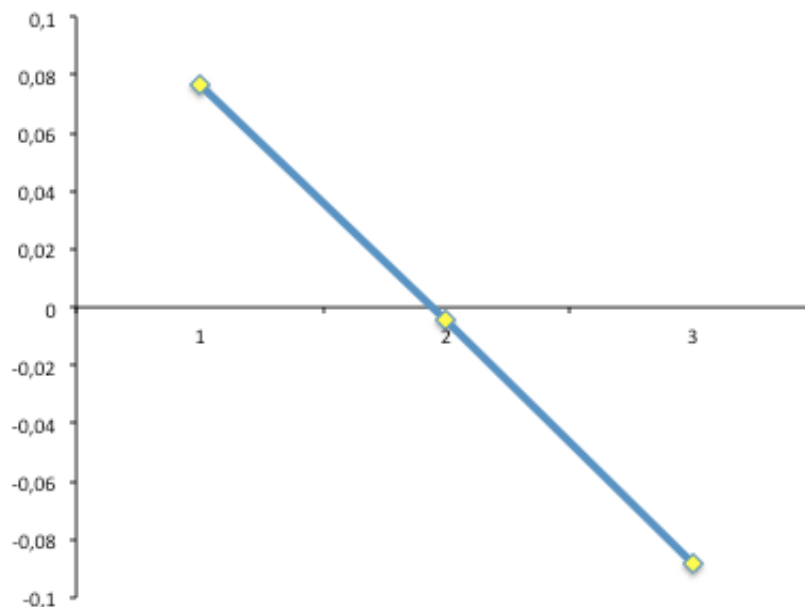
Paritet	Resultat	Obs	Std. Dev.	Percent
1	0.0765832	20564	0.9840825	39.12
2	-0.0041045	19293	0.9893676	36.70
3	-0.0882931	9477	1.020792	18.03
4	-0.1745895	2394	1.032313	4.55
5	-0.3103126	595	1.081022	1.13
6	-0.1992412	122	1.069189	0.23
7	-0.3986116	54	1.239992	0.10
8	-0.2544794	37	1.165278	0.07
9	-0.1376593	13	1.01724	0.02
10	-0.4960847	11	1.35791	0.02
11	0.3014593	7	1.054709	0.01
12	0.4099351	3	0.8219998	0.01
13	0.9477945	3	0.6175645	0.01

Av tabellen ser vi at resultatet faller med stigende søskenparitet - i alle fall for de lavere paritetene. Resultatene går jevnt nedover helt til femtefødte i en søskenflokk. Deretter går de litt sporadisk opp og ned, men her er det veldig få observasjoner og vi vil dermed her ikke få virkelig gode estimater. En artig observasjon er de tre registrerte elevene som er født som nummer 13 i søskenflokk, for vi ser av tabellen at disse presterer best i testen med klar margin.

Vi har samlet resultatene til første-, andre- og tredjefødte i figur 3.2 der vi har *Resultat* på

Y-aksen og søskenrekkefølgen, *Paritet*, på X-aksen. Vi ser her tydelig dårligere resultater på elever som er født lenger ut i søskenflokket.

Figur 3.2: Avhengig- og interessevariabel



### 3.2.3 Kontrollvariable

Det er viktig å ha med kontrollvariable i begge analysene. Disse inkluderes for å utelukke at sammenhengen mellom interessevariabelen og den avhengige variabelen skyldes andre variabler. I våre analyser er det tydelig flere effekter som påvirker resultatet på nasjonale prøver enn kun søskenrekkefølgen. Det er derfor viktig for oss å ha med kontrollvariabler i begge regresjonene. I første regresjon for å skille ut effekten av søskenrekkefølgen, og i andre regresjon for å få mest mulig korrekt estimat på skoleeffektene. Dette kommer vi tilbake til i neste kapittel.

Kontrollvariablene vi inkluderer er den individspesifikke variabelen for kjønn, de familie-spesifikke variablene for etnisk bakgrunn, mor- og fars utdanning og inntekt og til slutt familiestørrelsen. Alle kontrollvariablene er vist i tabell 3.3.

Kjønn karakteriseres ved en dummy-variabel lik 1 dersom eleven er gutt og 0 dersom eleven er jente. Datasettet inneholder 49% gutter.

Tabell 3.3: Data for kontrollvariabler

Variable	Obs.	Mean	Std. Dev.	Min	Max
Kjønn	52573	0.4904609	0.4999138	0	1
Innvandrerbakgrunn	52573	0.1252354	0.3309888	0	1
Mor utdanning	52573	4.406787	1.625419	1	8
Far utdanning	52573	4.319308	1.62487	1	8
Mor inntekt	52573	3.623064	3.92408	1	20
Far inntekt	52573	4.184194	3.284324	1	20
Antall søsken	52573	1.832062	1.128914	0	15

Vi har generert en dummy der  $Innvandrer=1$  betyr at eleven har enten mor eller far med utenlandsk opprinnelse og  $Innvandrer=0$  betyr at begge foreldre er etnisk norske. Av tabellen ser vi at ca 12,5% har utenlandsk opprinnelse.

Videre har vi generert en variabel for mors og fars utdanning som er henholdsvis *Mor utdanning* og *Far utdanning*. Disse variablene varierer fra 1-8 der 1 er lite eller ingen utdanning, og 8 er høyest mulig utdanningsnivå som tilsvarer doktorgrad/forskerutdanning. Viser tallenes betydning i tabell 3.4 som vi har hentet fra statistisk sentralbyrå (Rognan & Barrabés 2001, p. 7)

Tabell 3.4: Utdanningstabell

Tredeling av nivå	Nivå	Nivånavn	Klasstrinn
	<b>0</b>	Ingen utdanning eller førskoleutdanning	Under skoleplikt
Obligatorisk utdanning	<b>1</b>	Barneskoleutdanning	1.-7. klasstrinn
	<b>2</b>	Ungdomsskoleutdanning	8.-10. klasstrinn
Mellomutdanning	<b>3</b>	Videregående, grunnutdanning	11.-12. klasstrinn
	<b>4</b>	Videregående, avsluttende utdanning	13. klasstrinn +
	<b>5</b>	Påbygging til videregående utdanning	14. klasstrinn +
Universitets- og høyskoleutdanning	<b>6</b>	Universitets- og høyskoleutdanning, lavere nivå	14.-17. klasstrinn
	<b>7</b>	Universitets- og høyskoleutdanning, høyere nivå	18.-19. klasstrinn
	<b>8</b>	Forskerutdanning	20. klasstrinn +
	<b>9</b>	Uoppgitt	

Utdanningsnivået for mor er i gjennomsnittet 4,41 og 4,32 for far. Vi ser av tabellen at dette tilsvarer fullført videregående for begge parter. Mor har noe høyere utdanningsnivå ifølge våre observasjoner, men det er ingen stor forskjell.

For mors og fars inntekt har vi ikke med ekstremverdier slik at maksimal inntekt er 2 000 000 og minimumsinntekten vi har tatt med er 100 000. Dette ser vi i tabell 3.3 der mini-

mumsverdien er 1 og maksimum er 20. Her ser vi at far i gjennomsnitt tjener 418 000 og mor 362 000.

I tabell 3.5 ser man fordelingen av elevenes familiestørrelse der vi har med variabelen for antall søsken. Den viser oss at de fleste elevene har enten ett eller to søsken.

Tabell 3.5: Antall søsken

Antall søsken	Freq.	Percent
0	2.394	4.55
1	20.597	39.18
2	19.148	36.42
3	6.772	12.88
4	2.310	4.39
5	812	1.54
6	277	0.53
7	116	0.22
8	70	0.13
9	37	0.07
10	19	0.04
11	11	0.02
12	3	0.01
13	5	0.01
14	1	0.00
15	1	0.00



# Kapittel 4

## Økonometrisk tilrettelegging

### 4.1 Kvantitativ metode

Vårt primære formål er å gjennomføre en kvantitativ analyse for å studere sammenhengen mellom skolekvalitet og søskenrekkefølgen. Vi ønsker å legge et grunnlag for denne analysen ved først å se om vi har evidens for søskenrekkefølgeeffekt i datasettet vårt, deretter se om skoleeffektene varierer mer for elever med høyere søskenparitet.

Den vanligste formen for kvantitativ analyse er regresjonsanalyse og denne skal vi benytte i vår oppgave. Det er en statistisk metode som undersøker sammenhengen mellom to eller flere variabler. Fordelen med en slik analyse er at den tar hensyn til kontrollvariabler som inkluderes i en regresjonsanalyse. Formålet er å utelukke at sammenhengen mellom de uavhengige og avhengige variablene ikke skyldes kontrollvariablene. Om det viser seg at sammenhengen mellom den avhengige og den uavhengige variabelen skyldes en eller flere kontrollvariabler, har vi en spuriøs sammenheng. Da er det en bakenforliggende variabel (en kontrollvariabel som vi ikke har med) som er årsaken til sammenhengen.

## 4.2 Regresjon

### 4.2.1 Første regresjon

I den første delen av analysen tar vi for oss vanlig MKM-regresjon, minste kvadraters metode, for å finne ut om søskenrekkefølgeeffekten er negativ som tidligere antatt. MKM-regresjonen lager en lineær linje for observasjonene slik at vi kan se hvordan variablene påvirker hverandre. Slik ligningen er spesifisert ligger alle skolevariabler i restleddet. Her antar vi, i tråd med flertallet av eksisterende empiriske analyser av søskenrekkefølgeeffekt, at søskenparitet ikke korrelerer med noen skolevariabler. Vi har derfor bare med individ- og familievariabler. Ligningen ser slik ut:

$$Resultat_i = \alpha_0 + \alpha_1 Paritet_i + \alpha_2 Individ_i + \alpha_3 Familie_i + \epsilon_i \quad (4.1)$$

Her er  $Resultat_i$  elevprestasjonen for individ  $i$ ,  $Paritet_i$  er interessevariabelen og  $i$  tillegg må vi kontrollere for individuelle trekk og familietrekk. Fullt utskrevet vil ligningen se ut som følger:

$$\begin{aligned} Resultat_{ij} = & \beta_0 + \beta_1 Paritet_i + \beta_2 Kjonn_i + \beta_3 Innvandreri + \beta_4 Mor.utdanning_i \\ & + \beta_5 Far.utdanning_i + \beta_6 Mor.inntekt_i + \beta_7 Far.inntekt_i + \beta_8 Ant.sosken_i + \epsilon_{ij} \end{aligned} \quad (4.2)$$

Mange tidligere studier har forsøkt å separere søskenrekkefølgeeffekten fra familiestørrelseeffekten. Litteraturen viser oss at disse to effektene ofte blir blandet med hverandre hvis vi ikke kontrollerer for det (Black et al. 2005). Vi innser imidlertid at familiestørrelsevariabelen,  $Ant.sosken_i$ , medfører endogenitetsproblemer, det vil si at familiestørrelsen reflekterer uobserverte egenskaper ved familieeffektene. Hvilke faktorer som fører til at noen familier er store, mens andre er små kan være mange. Vi vet at folk med lav utdanning og lav inntekt ofte har flere barn enn de med høy utdanning og inntekt, men vi vet lite om hvilke andre faktorer som spiller inn. I litteraturen kan vi se at det er vanlig å bruke instrumentvariabel-

metoden (IV) for å kontrollere for endogenitet når familiestørrelsevariabelen blir brukt. Her blir et instrument, for eksempel tvillinger, brukt til å estimere den kausale effekten slik at vi kan få konsistente estimater selv om familiestørrelsevariabelen er korrelert med restleddet. Vi bruker ikke IV-metoden for å takle endogenitetsproblemet, men vi innser at det er viktig å ha med familiestørrelsevariabelen fordi vi mistenker at mye av effekten til *Paritet*-variabelen kommer fra familiestørrelse. Hvis eleven har høy søskenparitet er også familien stor, men det er vanskelig å skille de to effektene fra hverandre.

I denne oppgaven hvor vi kun ser på et tverrsnitt, mister vi muligheten til å se på "within"-variasjonen i familiene. For å kunne undersøke forskjellene mellom søsken innad i familien er vi avhengig av å ha observasjoner på flere tidspunkter, men i vår tverrsnittanalyse har vi kun ett tidspunkt og må derfor sammenligne elever fra forskjellige familier. Det kan være problematisk å sammenligne elever med lik søskenparitet fra forskjellige familier, siden bidraget de har med seg fra hjemmet kan variere stort.

Videre i første regresjon estimerer vi en spesifisering med dummyvariabler for første-, andre- og tredjefødt (*Par1*, *Par2* og *Par3*) for å kontrollere for søskenrekkefølgeeffekten for de forskjellige søskenparitetene.

$$\begin{aligned} \text{Resultat}_{ij} = & \beta_0 + \beta_1 \text{Par2} + \beta_2 \text{Par3} + \beta_3 \text{Kjonn}_i + \beta_4 \text{Innvandrer}_i + \beta_5 \text{Mor.utdanning}_i \\ & + \beta_6 \text{Far.utdanning}_i + \beta_7 \text{Mor.inntekt}_i + \beta_8 \text{Far.inntekt}_i + \beta_9 \text{Ant.soksen}_i + \epsilon_{ij} \end{aligned} \quad (4.3)$$

Ved å bruke *Par1* som referansekategori vil vi observere forholdet mellom andre- eller tredjefødt relativt til førstefødt, og størrelsen på effektene<sup>1</sup>. Ifølge den første hypotesen vi tester ut vil begge disse effektene være negative, og hvis litteraturen som mener at den negative effekten øker med søskenpariteten medfører riktighet vil *Par3* være mer negativ enn *Par2*

---

<sup>1</sup>Se ligning (4.3)

### 4.2.2 Andre regresjon

I og med at vi kun har individ- og familiespesifikke variabler i produktfunksjonen vår blir skoleeffektene fanget opp i restleddet. Vi ønsker å se på en ligning som ser slik ut:

$$Resultat_{ij} = \alpha_0 + \alpha_1 Individ_i + \alpha_2 Familie_i + \alpha_3 Skole_j + \epsilon_i \quad (4.4)$$

Vi legger her på en variabel,  $Skole_j$ , som er effekten skole  $j$  har på elevprestasjonen til individ  $i$ . Gjennomsnittsskolen vil ha effekten tilsvarende verdien til konstantleddet på elevprestasjon, fordi den skolefaste effekten vil være null. Ved å skille ut skoleeffektene fra restleddet ser vi hvor mye en gitt skole avviker fra gjennomsnittsskolen når det gjelder innflytelse på elevprestasjon. Legg merke til at vi her tar bort paritetsvariablene, dette fordi vi ønsker å se "whitin"-variasjonen for skoleeffekt for hver av søskenparitetene, og derfor må vi estimere skoleeffektene for første-, andre- og tredjefødte hver for seg. På denne måten kan vi analysere variasjonen i skoleeffekt for de tre søskenparitetene adskilt. For å finne de skolefaste effektene for hver søskenparitet erstattes restleddet med:

$$\epsilon_{ij} = v_j + \eta_{ij} \quad (4.5)$$

Ender opp med:

$$Resultat_{ij} = \alpha_0 + \alpha_1 Individ_i + \alpha_2 Familie_i + v_j + \eta_{ij} \quad (4.6)$$

Fullt utskrevet med skolefaste effekter blir ligningen seende ut som følger:

$$Resultat_{ij} = \beta_0 + \beta_1 Kjonni + \beta_2 Innvandrer_i + \beta_3 Mor.utdanning_i + \beta_4 Far.utdanning_i + \beta_5 Mor.inntekt_i + \beta_6 Far.inntekt_i + \beta_7 Ant.soksen_i + v_j + \eta_{ij} \quad (4.7)$$

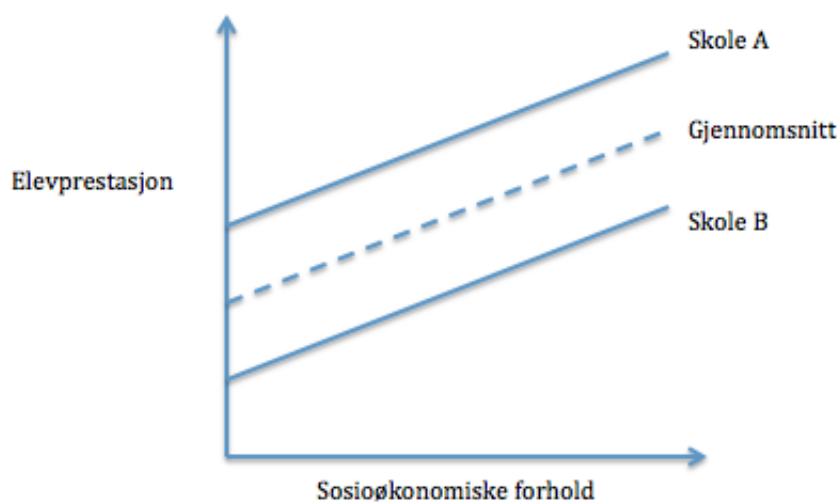
Skoleeffekten fanger opp alle forhold som kan gjøre at en skole er bedre eller dårligere enn

en annen, som for eksempel lærertetthet, klassestørrelser og ressurser. Ved å anvende faste effekter-metoden deler vi restleddet opp i to, ett skolespesifikt restledd,  $v_j$ , og ett restledd som fanger opp resten av uobserverte effekter,  $\eta_{ij}$ <sup>2</sup>.

Vanligvis brukes faste effekter i paneldataanalyser til å fjerne individspesifikke effekter som ikke varierer over tid. Vi ser her kun på ett år og ett bestemt klassetrinn, så dette er ikke vårt fokus. De skolefaste effektene kan tolkes som den samlede gjennomsnittsinflytelsen skolen har på elevens prestasjon. Vår interesse er derfor knyttet til variasjonen i de skolefaste effektene.

Som vi ser av figur 4.1 vil vi ha to skoler med ulik skoleeffekt,  $v_j$ . Derfor kan to elever fra ulike skoler, men med like sosioøkonomiske forhold oppleve forskjellige resultater,  $Resultat_{ij}$ . Vi ser på figuren at gjennomsnittsskolen ligger mellom skole A og skole B (stiplet linje), og dette betyr at skole A har positiv effekt, mens skole B har negativ effekt på prestasjon relativt til gjennomsnittsskolen.

Figur 4.1: Skoleeffekt



Skjæringspunktet på grafen bestemmes av konstantleddet,  $\alpha_0$ , + restleddet,  $v_j$ <sup>3</sup>. Sosioøkonomiske fordeler forbedrer prestasjonene og er derfor en positiv sammenheng. Hypotesen sier

<sup>2</sup>Se ligning (4.5)

<sup>3</sup>Se ligning (4.6)

at variansen til skoleeffektene er større for elever med høy søskenparitet. Vi estimerer derfor ligning (4.7) separat for første-, andre- og tredjefødte.

Videre i andre regresjon undersøker vi hvor sterkt skoleeffektene for de tre søskenparitetene korrelerer. Dette vil være en indikasjon på om skoler som er gode for én søskenparitet også er gode for andre søskenpariteter. Dette gjøres ved å sette opp en korrelasjonsmatrise hvor vi vil observere hvor mye/lite skoleeffektene korrelerer med hverandre. Vi vil også estimere ligning (4.8) for å se hvordan en økning i skoleeffekten for andre- og tredjefødte påvirker førstefødte.

$$Skoleeffekt1 = \alpha_0 + \alpha_1 Skoleeffekt_i + \eta, i = 2, 3 \quad (4.8)$$

# Kapittel 5

## Resultater

### 5.1 Tverrsnittanalyse

Estimerer nå sammenhengen mellom avhengig variabel, elevprestasjon og interessevariabelen søskenparitet. For å unngå skjevhet blir kontrollvariablene inkludert. Videre estimerer vi de skolefaste effektene og ser på variasjonen for hver enkelt søskenparitet.

### 5.2 Første regresjonsmodell

Først ønsker vi å bekrefte at vi har evidens for søskenrekkefølgeeffekt i dataen. Anvender derfor MKM-regresjon på ligning (4.2). Her er *Paritet* uavhengig variabel, og kontrollvariablene i ligningen inkluderes suksessivt.

Resultatene er gjengitt i tabell 5.1.

Tabell 5.1: Interesse- og kontrollvariabler

Variabler	1	2	3
<b>Paritet</b>	<b>-0.07804</b>	<b>-0.0572802</b>	<b>-0.0373914</b>
<i>Std. Err.</i>	<i>0.004373</i>	<i>0.0054473</i>	<i>0.005172</i>
<b>Antall søsken</b>		<b>-0.0306452</b>	<b>-0.0113791</b>
<i>Std. Err.</i>		<i>0.0047979</i>	<i>0.0045574</i>
<b>Kjønn</b>			<b>-0.0707329</b>
<i>Std. Err.</i>			<i>0.0082269</i>
<b>Innvandrerbakgrunn</b>			<b>-0.0326189</b>
<i>Std. Err.</i>			<i>0.0124839</i>
<b>Mor utdanning</b>			<b>0.1216291</b>
<i>Std. Err.</i>			<i>0.0028495</i>
<b>Mor inntekt</b>			<b>0.0017615</b>
<i>Std. Err.</i>			<i>0.0010605</i>
<b>Far utdanning</b>			<b>0.1079002</b>
<i>Std. Err.</i>			<i>0.0028517</i>
<b>Far inntekt</b>			<b>0.0077605</b>
<i>Std. Err.</i>			<i>0.0012737</i>
<b>Konstant</b>	<b>0.1512869</b>	<b>0.1671861</b>	<b>-0.9087911</b>
<i>Std. Err.</i>	<i>0.0095275</i>	<i>0.0098438</i>	<i>0.0179046</i>
$R^2$	<b>0.0060</b>	<b>0.0068</b>	<b>0.1110</b>
<b>Antall observasjoner</b>	<b>52573</b>	<b>52573</b>	<b>52573</b>

I kolonne én hvor vi kun har interessevariabelen *Paritet*, avtar resultatene for høyere søskenparitet. Øker vi søskenpariteten med én, for eksempel andrefødt og ikke førstefødt, vil resultatet gå ned med 0,07804 (7,8%). Er du tredjefødt vil resultatet avta med  $0,07804 \cdot 2$ , fjerdefødt  $0,07804 \cdot 3$  osv.

I kolonne to inkluderes en ekstra variabel, *Antall søsken*, som kontrollerer for familiestørrelsen. Estimaten viser en negativ familiestørrelseseffekt, som tar bort en del av søskenrekkefølgeeffekten. Det vil si at større familier betyr dårligere resultater for eleven. Vi har fortsatt en negativ sammenheng mellom søskenparitet og resultater, men estimaten er lavere enn før da vi nå har inkludert variabelen for familiestørrelseseffekten. Et ekstra familiemedlem betyr at resultatene for den enkelte elev reduseres med 0,0306. Med andre ord vil, for gitt



søskenparitet, prestasjonen reduseres med ca. 3% for hvert nytt søsken. Og for gitt familiestørrelse vil prestasjonene reduseres med 0,057 hvis søskenparitet øker med én. Dette er en dobbelt negativ effekt som i kolonne én bare fanges opp i *Paritet*-variabelen.

Disse funnene er like de Black et al. (2005) rapporterte i sin undersøkelse av søskenrekkefølge, familiestørrelse og intelligens. Undersøkelsen rapporterte først negativ familieeffekt og observerte at den ble redusert når søskenrekkefølgeeffekten ble tatt hensyn til. Vi har gått motsatt vei ved å først finne negativ effekt forbundet med søskenparitet, slik at søskenrekkefølgeeffekten ble redusert da vi tok hensyn til familiestørrelse. Begge metodene viser at disse to effektene henger sammen og den ene må kontrolleres for hvis vi vil studere den andre.

I kolonne tre, i tabell 5.1, inkluderes de resterende seks kontrollvariablene. Selv når vi inkluderer disse kontrollvariablene ser vi en tydelig negativ sammenheng mellom høyere søskenparitet og resultatet på den nasjonale prøven. Effekten blir lavere med flere kontrollvariable inkludert i regresjonen, men effekten er uansett statistisk signifikant og av en viss størrelse. Vi har nå evidens for vår første hypotese der høyere søskenparitet betyr dårligere resultater. Øker vi søskenpariteten med én vil resultatet reduseres med 0,037. Som før er variabelen for familiestørrelsen fortsatt negativ, men den negative effekten er ikke like stor som før. Nå betyr et ekstra familiemedlem en reduksjon med 0,011 på resultatet, noe som er en tredjedel av hva som var tilfelle i regresjonen med kun denne variabelen og søskenpariteten (kolonne to). En mulig tolkning av dette er at familiestørrelsen korrelerer med foreldrenes utdanningsnivå, slik at punktestimatet for familiestørrelse reduseres når disse kontrollvariablene inkluderes. Med andre ord; høyt utdannede foreldre (spesielt mor) betyr ofte mindre familier, der de kan gi barnet mer ressurser (tid og penger), relativt til foreldre med lavere utdanning (McLanahan 2004).

Av variabelen kjønn ser vi at resultatet reduseres med 0,0707 når eleven er gutt. Estimaten viser at jentene i dette kullet presterer noe bedre enn guttene. Har eleven enten mor eller far med utenlandsk opprinnelse er resultatene 0,0326 dårligere enn de elevene med norske foreldre. Variablene for mors og fars inntekt og for mors og fars utdanning har et positivt fortegn. Det er dermed en positiv sammenheng mellom disse variablene og resultatet til eleven. Høyere inntekt og utdanning hos foreldrene betyr bedre resultat for eleven. Øker vi

utdanningsnivået til mor og far med én, øker resultatene med henholdsvis 0,122 og 0,108 for eleven. Som vi så i tabell 3.4 fra statistisk sentralbyrå (Rognan & Barrabés 2001) tilbake i kapittel 3, vil det å øke utdanningsnivå med én, for eksempel fra 6 til 7, bety at foreldrenes utdanningsnivå øker fra bachelorgrad til mastergrad. Da vil eleven med foreldre som har fullført mastergrad ha  $0,122+0,108=0,230$  (23%) bedre resultat enn eleven med foreldre med fullført bachelorgrad<sup>1</sup>. Denne økningen er lik for hvert eneste utdanningsnivå helt fram til siste som er 8 og betyr foreldre med utdanningsnivå doktorgrad eller mer. Øker vi inntekten til mor med 100 000, øker resultatet til eleven med 0,0017. For far betyr økt inntekt med 100 000 resultatforbedring til eleven med 0,0076. En økning i foreldrenes inntekt har dermed ikke mye å si for prestasjonen til elevene.

### 5.2.1 T-test

For å bekrefte at vi har negativ søskenrekkefølgeeffekt må vi teste om koeffisienten til *Paritet*<sup>2</sup> er signifikant forskjellig fra 0. Dette gjøres lett ved en t-test. Her lager vi oss en hypotese,  $H_0$ , som sier at  $\hat{\beta}_1 = 0$ . Alternativhypotesen,  $H_1$  sier at  $\hat{\beta}_1 \neq 0$ . Så finner vi  $t_{score}$  og undersøker om absoluttverdien er større, mindre eller lik  $t_{value}$ .

$$t_{score} = \frac{\hat{\beta} - \beta_1}{Std.Err_{\hat{\beta}}} \sim \tau_{n-k-1} \quad (5.1)$$

I vårt tilfelle har vi:

$$t_{score} = \frac{-0,0373914}{0,005172}^3 = -7,2296 \quad (5.2)$$

Vi finner kritisk verdi ved å ta antall observasjoner minus antall variabler minus én,  $\tau_{n-k-1}$ . Da har vi funnet antall frihetsgrader og kan lese kritisk verdi ut av tabellen *Critical Values of the t Distribution* (Wooldridge 2012, p. 825) med signifikansnivå 95%, og den kritiske

<sup>1</sup>Antar lineær sammenheng mellom alle utdanningsnivåene, selv om vi mistenker avtagende effekt ved høyere utdanningsnivåer

<sup>2</sup>Se ligning (4.2)

<sup>3</sup>Se tabell 5.1

verdien blir da 1,96. Vi ser at  $|-7,2296| > 1,96$ , og vi forkaster  $H_0$ . Det vil si; koeffisienten til *Paritet* er signifikant forskjellig fra null og den negative sammenhengen mellom søskenparitet og elevprestasjon er dermed til stede i vårt datasett. Resultatene fra denne delen av analysen er i god overensstemmelse med den eksisterende empiriske litteratur om betydning av søskenrekkefølge.

### 5.2.2 Oppdeling av paritetsvariabelen

Som forklart i kapittel 3 og 4 deler vi opp paritetsvariabelen i tre: *Par1* betyr førstefødt, *Par2* er andrefødt og *Par3* er tredjefødt. Vi skal nå vise hvordan alle variablene reagerer med førstefødt, *Par1*, som referansevariabel. Ser resultatene i tabell 5.2.

I kolonne én har vi kun med variablene for andre- og tredjefødte. Resultatet vil fortsatt gå ned ved høyere søskenparitet. Andrefødt gjør det 0,043 dårligere enn førstefødte, og tredjefødte gjør det atskillig dårligere, 0,127, enn førstefødte.

Som i tabell 5.1 legger vi først på kontrollvariabelen for familiestørrelsen *Antall søsken* i kolonne to. Det er fortsatt en negativ sammenheng mellom resultater og familiestørrelse. For hvert ekstra familiemedlem vi inkluderer vil førstefødte gjøre det 0,0556 dårligere på den nasjonale prøven. Ved å inkludere familiestørrelseeffekten reduseres den negative effekten til tredjefødte i stor grad, men ikke for andrefødte. *Par3* endres her fra -0,127 til -0,0881 slik at søskenrekkefølgeeffekten for tredjefødte ikke er like sterk som før, mens for andrefødte er effekten noenlunde lik. Variabelen for tredjefødte, *Par3*, fikk dermed ekstra negativ effekt i kolonne én fordi den inkluderte den negative effekten familiestørrelsen (som vi nå har skilt ut) har på resultatet.

I kolonne tre har vi inkludert resterende kontrollvariabler som i tabell 5.2, og selve estimatene på kontrollvariablene er nokså like som i tabell 5.1, og forklaringen er dermed lik. Det interessante her er estimatene for *Par2* og *Par3*. Gapet mellom *Par1*, som er referansevariabel, og *Par2* er mindre enn gapet mellom *Par2* og *Par3*. Tredjefødte gjør det derfor relativt dårligere i forhold til førstefødt enn andrefødte gjør i forhold til førstefødte. Prestasjonene avtar dermed ikke lineært, men det er fortsatt en klar sammenheng mellom dårligere prestasjoner

Tabell 5.2: Interesse- og kontrollvariabler

Variabler	1	2	3
<b>Par2</b>	<b>-0.0425846</b>	<b>-0.0467462</b>	<b>-0.0346418</b>
<i>Std. Err.</i>	<i>0.0096774</i>	<i>0.0096641</i>	<i>0.0091452</i>
<b>Par3</b>	<b>-0.1267732</b>	<b>-0.0880613</b>	<b>-0.0753052</b>
<i>Std. Err.</i>	<i>0.0121338</i>	<i>0.0124239</i>	<i>0.011764</i>
<b>Antall søsken</b>		<b>-0.0556269</b>	<b>-0.0262558</b>
<i>Std. Err.</i>		<i>0.0039788</i>	<i>0.0037926</i>
<b>Kjønn</b>			<b>-0.0708127</b>
<i>Std. Err.</i>			<i>0.0082286</i>
<b>Innvandrerbakgrunn</b>			<b>-0.0341962</b>
<i>Std. Err.</i>			<i>0.0124878</i>
<b>Mor utdanning</b>			<b>0.1225239</b>
<i>Std. Err.</i>			<i>0.002843</i>
<b>Mor inntekt</b>			<b>0.0018002</b>
<i>Std. Err.</i>			<i>0.0010606</i>
<b>Far utdanning</b>			<b>0.1077279</b>
<i>Std. Err.</i>			<i>0.0028515</i>
<b>Far inntekt</b>			<b>0.0077192</b>
<i>Std. Err.</i>			<i>0.0012738</i>
<b>Konstant</b>	<b>0.0384801</b>	<b>0.1349409</b>	<b>-0.9306645</b>
<i>Std. Err.</i>	<i>0.006475</i>	<i>0.0094538</i>	<i>0.0175421</i>
$R^2$	<b>0.0021</b>	<b>0.0058</b>	<b>0.1108</b>
<b>Antall observasjoner</b>	<b>52573</b>	<b>52573</b>	<b>52573</b>

for økt søskenparitet.

### 5.2.3 Skoleeffekter

I lys av resultatene fra første regresjonsmodell fortsetter vi med å undersøke skoleeffektene. Vi gjennomfører samme regresjon, men med faste effekter-metoden på skoleidentifikasjonen. Vi undersøker først om punkttestimatet for *Paritet* påvirkes ved inkludering av skolefaste effekter, deretter inkluderer vi resterende kontrollvariabler. Resultatene gjengis i tabell 5.3.

Tabell 5.3: Interesse- og kontrollvariabler m/skoleeffekt

Variabler	1	2	3
<b>Par2</b>	<b>-0.0376115</b>	<b>-0.0402505</b>	<b>-0.0293267</b>
<i>Std. Err.</i>	<i>0.0094421</i>	<i>0.0094369</i>	<i>0.0090238</i>
<b>Par3</b>	<b>-0.1089648</b>	<b>-0.0828106</b>	<b>-0.0724557</b>
<i>Std. Err.</i>	<i>0.0119106</i>	<i>0.0121935</i>	<i>0.0116661</i>
<b>Antall søsken</b>		<b>-0.0391373</b>	<b>-0.0193015</b>
<i>Std. Err.</i>		<i>0.0039837</i>	<i>0.0038324</i>
<b>Kjønn</b>			<b>-0.0777574</b>
<i>Std. Err.</i>			<i>0.0081214</i>
<b>Innvandrerbakgrunn</b>			<b>-0.0710822</b>
<i>Std. Err.</i>			<i>0.012928</i>
<b>Mor utdanning</b>			<b>0.1163809</b>
<i>Std. Err.</i>			<i>0.0028649</i>
<b>Mor inntekt</b>			<b>0.000146</b>
<i>Std. Err.</i>			<i>0.0010552</i>
<b>Far utdanning</b>			<b>0.0977475</b>
<i>Std. Err.</i>			<i>0.0028782</i>
<b>Far inntekt</b>			<b>0.0035892</b>
<i>Std. Err.</i>			<i>0.0012794</i>
<b>Konstant</b>	<b>0.0334449</b>	<b>0.1014007</b>	<b>-0.8443908</b>
<i>Std. Err.</i>	<i>0.0062555</i>	<i>0.0093222</i>	<i>0.0181682</i>
$R^2$	<b>0.0021</b>	<b>0.0057</b>	<b>0.1103</b>
<b>Antall observasjoner</b>	<b>52573</b>	<b>52573</b>	<b>52573</b>

### Utelatt variabelskjevhet

Ved å inkludere skolefaste effekter observerer vi at de fleste effektene avtar. Vi ønsker å fokusere på variablene for andre- og tredjefødte, *Par2* og *Par3*, og ser at effekten går ned for begge variablene når vi inkluderer skoleeffekter. Søskenrekkefølgeeffekten blir dermed mindre enn tidligere. Tolkningen av dette kan være at spesifikasjonen uten skoleestimer gir punkttestimer som er skjeve grunnet utelatte skolevariabler<sup>4</sup>. Da er det en bakenforliggende årsak som påvirker variabelen vi undersøker, og denne årsaken fanges opp i restleddet slik at

<sup>4</sup>Omitted variable bias problem

variabelen<sup>5</sup> korrelerer med restleddet. I spesifikasjonen uten skoleeffekter ligger skolevariablene i restleddet, og observerer at når vi inkluderer faste skoleeffekter i andre spesifikasjon har disse skolevariablene betydning for søskenrekkefølgen siden estimatene endrer seg. I regresjonen med faste skoleeffekter er mye av skoleinnflytelsen på prestasjonene samlet i den faste effekten og korrelasjonen mellom søskenparitetsvariablene og restleddet er nå blitt mindre. En interessant observasjon er at selv om søskenrekkefølgeeffekten på resultater har blitt redusert, er det relative gapet mellom *Par2* og *Par3* blitt større. Det betyr at de skolefaste effektene bidrar til større relative forskjeller mellom andre- og tredjefødte.

Av de andre familiespesifikke variablene reduseres også estimatene for antall søsken og foreldrenes inntekt og utdanning. Variablene har nå mindre påvirkning på elevprestasjon når skolefaste effekter legges til. Den individspesifikke variabelen *Kjønn* øker marginalt, som betyr at jentene gjør det marginalt bedre enn før.

Vi legger merke til at effekten for den familiespesifikke variabelen, *Innvandrer*, får en økt negativ effekt ved å inkludere skolefaste effekter. På samme måte som i resonnetet for søskenparitetene kan årsaken være at innvandrervariabelen er korrelert med restleddet. Vi kan tenke oss en viktig skolevariabel som lærerkvalitet. Effekten av lærerkvalitet på prestasjoner antas å være positiv, slik at prestasjoner hos elever øker for økt lærerkvalitet. Dermed er korrelasjonen mellom lærerkvalitet og innvandrers skoleprestasjon også positiv. I første spesifikasjon uten skoleeffekter tar ikke regresjonen hensyn til dette og denne effekten ligger i variabelen for innvandrere, men ved å inkludere skoleeffekter er denne effekten "skilt ut" fra innvandrervariabelen. I og med vi har positiv korrelasjon mellom skolevariablene og innvandrervariabelen har vi en positiv skjevhet i spesifikasjonen uten faste skoleeffekter.

For å oppsummere vil korrelasjonen mellom variablene og restleddet bli mindre ved inkludering av skolefaste effekter. Vi har nå fått mer presise estimater da vi har redusert korrelasjonen mellom variablene og restleddet. Absoluttverdien til de estimatene som viser familievariablenes påvirkning blir redusert foruten variabelen for innvandrer, mens estimatene for den individuelle variabelen for kjønn øker marginalt. Variablenes fortegn er det samme, slik at det for eksempel teller positivt på elevens prestasjon med høy utdanning hos mor og

---

<sup>5</sup>Her: søskenparitet

far.

### Skoleeffekt for den enkelte søskenparitet

Vi gjennomfører igjen en regresjon, men nå for å se hvordan skoleeffekten påvirker den enkelte søskenparitet. Vi estimerer ligning (4.7).

Tabell 5.4 viser estimater for alle variablene for de ulike søskenparitetene. I kolonne én har vi kun estimater for førstefødte, i andre kolonne for andrefødte og for tredjefødte i kolonne tre.

Tabell 5.4: Variabler inkl. skoleeffekter

Paritet	1	2	3
<b>Antall søsken</b>	<b>0.008268</b>	<b>-0.0041866</b>	<b>-0.0536243</b>
<i>Std. Err.</i>	<i>0.0075118</i>	<i>0.0083017</i>	<i>0.013527</i>
<b>Kjønn</b>	<b>-0.06901</b>	<b>-0.0976081</b>	<b>-0.0619686</b>
<i>Std. Err.</i>	<i>0.0132084</i>	<i>0.0138682</i>	<i>0.0215456</i>
<b>Innvandrerbakgrunn</b>	<b>-0.0744673</b>	<b>-0.0716609</b>	<b>-0.0405487</b>
<i>Std. Err.</i>	<i>0.0206316</i>	<i>0.0222136</i>	<i>0.034897</i>
<b>Mor utdanning</b>	<b>0.1169845</b>	<b>0.1080939</b>	<b>0.1148994</b>
<i>Std. Err.</i>	<i>0.0047566</i>	<i>0.0048848</i>	<i>0.0074821</i>
<b>Mor inntekt</b>	<b>0.0004007</b>	<b>-0.0017262</b>	<b>0.0040703</b>
<i>Std. Err.</i>	<i>0.0016803</i>	<i>0.0018711</i>	<i>0.0029296</i>
<b>Far utdanning</b>	<b>0.104764</b>	<b>0.0945409</b>	<b>0.0943619</b>
<i>Std. Err.</i>	<i>0.0046451</i>	<i>0.0049293</i>	<i>0.0076795</i>
<b>Far inntekt</b>	<b>0.0030071</b>	<b>0.0040943</b>	<b>0.0060352</b>
<i>Std. Err.</i>	<i>0.0019818</i>	<i>0.0022972</i>	<i>0.0035336</i>
<b>Konstant</b>	<b>-0.8999843</b>	<b>-0.8357252</b>	<b>-0.8543801</b>
<i>Std. Err.</i>	<i>0.0280893</i>	<i>0.0306053</i>	<i>0.0549258</i>
$R^2$	<b>0.1135</b>	<b>0.0985</b>	<b>0.1071</b>
<b>Antall observasjoner</b>	<b>20564</b>	<b>19293</b>	<b>9477</b>

Sammenligner vi variabelen for familiestørrelseseffekten i tabell 5.4 med 5.3 ser vi en meget interessant observasjon. I tabell 5.4 er variabelen for familiestørrelsen *Antall søsken* ikke

signifikant for første- og andrefødte, men derimot for tredjefødte. Beviser dette av t-testene vi har foretatt for alle søskenparitetene i tabell 5.5.

Tabell 5.5: T-test for familiestørrelsen

Paritet	1	2	3
$t_{score}$	1.10	-0.50	-3.96
$t_{value}$	1.96	1.96	1.96
	Ikke Signifikant	Ikke Signifikant	Signifikant

For tredjefødte vil derfor resultatet reduseres med 0,053 (5,3%) per ekstra familiemedlem, men for første- og andrefødte har familiestørrelsen ingen effekt på resultatet. Dermed er den negative effekten vi får på familiestørrelseseffekten i tabell 5.3 (som er signifikant) kun drevet av effekten for tredjefødte, *Par3*.

Tabell 5.4 viser at variabelen for kjønn er mer negativ for andrefødte enn for første- og tredjefødte. Gutter gjør det dårligere enn jentene ved alle søskenparitetene, men forskjellen mellom kjønnene er klart størst for andrefødte. For første- og andrefødte vil elever med innvandrerbakgrunn gjøre det ca. 7% dårligere enn etnisk norske elever. Det interessante er at denne effekten er fraværende hos tredjefødte da denne ikke er signifikant. Tredjefødtes resultater påvirkes derfor ikke av etnisk bakgrunn. Av resterende variabler ser vi at det er omtrent like estimater ved alle søskenparitetene. Ingen av søskenparitetene påvirkes signifikant av mors og fars inntekt, mens de blir påvirket i noenlunde lik grad av mors og fars utdanning.



### 5.3 Andre regresjonsmodell

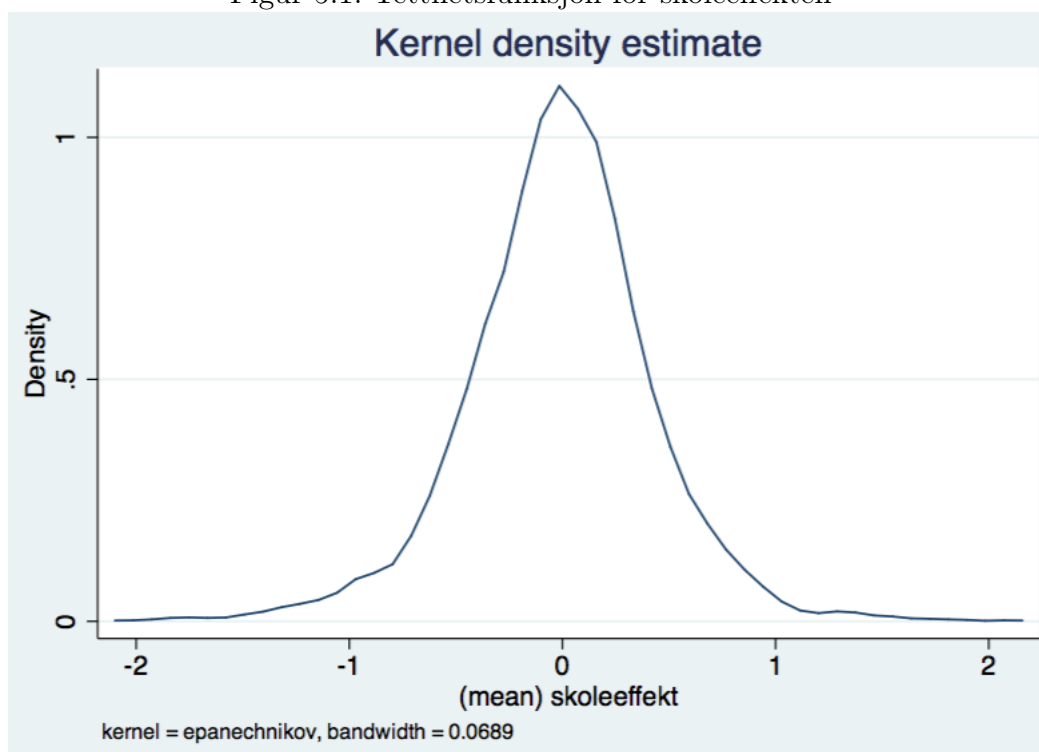
Nå går vi videre til hovedbidraget i oppgaven, nemlig hypotesen om at skolekvalitet har mer å si for senerefødte i en søskenflokk. For å undersøke dette tar vi utgangspunkt i estimatene fra tabell 5.3 der vi har inkludert skoleeffekter. Vi predikerer restleddet som transformeres bort,  $v_j$ , som i vårt tilfelle representerer skoleeffekten som varierer fra skole til skole. Da får vi informasjon om skoleeffekten til alle elevene. Viser resultatene i tabell 5.6.

Tabell 5.6: Skoleeffekt

Variabel	Antall obs	Mean	Std. Dev.	Min	Max
Skoleeffekt	2392	-0.0148492	0.4374013	-2.030926	2.088657

Ser av tabellen at alle 52 573 elevene vi har med i datasettet er fordelt på 2392 forskjellige skoler. Viser fordelingen av skoleeffekten grafisk i figur 5.1.

Figur 5.1: Tetthetsfunksjon for skoleeffekten



Her ser vi tetthetsfunksjonen for skoleeffekten til alle elevene. I vår oppgave ønsker vi å undersøke disse effektene for den enkelte søskenparitet. Vi får da testet hovedhypotesen om

at senerefødte påvirkes i større grad av skolekvaliteten enn tidligerefødte. Da tar vi utgangspunkt i den predikerte variabelen for skoleeffekten og endrer den slik at vi ser skoleeffekten kun for den enkelte søskenparitet. Vi får dermed sjekket hypotesen ved å se hvordan skoleeffektene varierer innad for hver enkelt søskenparitet. Som før har vi kun med tre søskenpariteter da disse utgjør ca. 94% av alle elevene. Effektene vises i tabell 5.7, der *Skoleeffekt1* er skolens påvirkning på førstefødt, *Skoleeffekt2* for andrefødte og *Skoleeffekt3* for tredjefødte.

Tabell 5.7: Skolefaste effekter

Variable	Antall obs.	Mean	Std. Dev	Min	Max
Skoleeffekt1	2210	-0.0295028	<b>0.5326607</b>	-2.398161	1.838276
Skoleeffekt2	2215	-0.0227875	<b>0.547524</b>	-2.339756	1.790441
Skoleeffekt3	2075	-0.0141238	<b>0.6550209</b>	-2.58347	2.224927

Av kolonne fire (Std. Dev) ser vi at resultatene til tredjefødte blir sterkere påvirket av skolekvaliteten enn for første- og andrefødte, noe som gjør at resultatet til tredjefødte varierer mer. Vi kan derfor tenke oss at dette har sammenheng med egenskapene til senerefødte som vi har lest i litteraturen. I modellen til Hotz & Pantano (2015) argumenteres det for at elevprestasjonen er et resultat av disiplinering i hjemmet. Vi har utvidet modellen der vi ser på elevprestasjon som et resultat av disiplinering i hjemmet og disiplinering fra skolen<sup>6</sup>. For et gitt resultat må dermed elever som har mangel på disiplin i hjemmet trenge mer disiplin fra skolen. Hotz & Pantano (2015) argumenterer for at familiedisiplineringen hjemme blir mildere når søskenrekkefølgen øker. Når disiplinen hjemme er i avtakende grad for søskenpariteten vil senerefødte ha behov for mer disiplin og oppfølging fra skolen enn tidligerefødte. Analysen vår underbygger dette da resultatene til tredjefødte avhenger i større grad av skolekvaliteten enn hva den gjør for første- og andrefødte<sup>7</sup>.

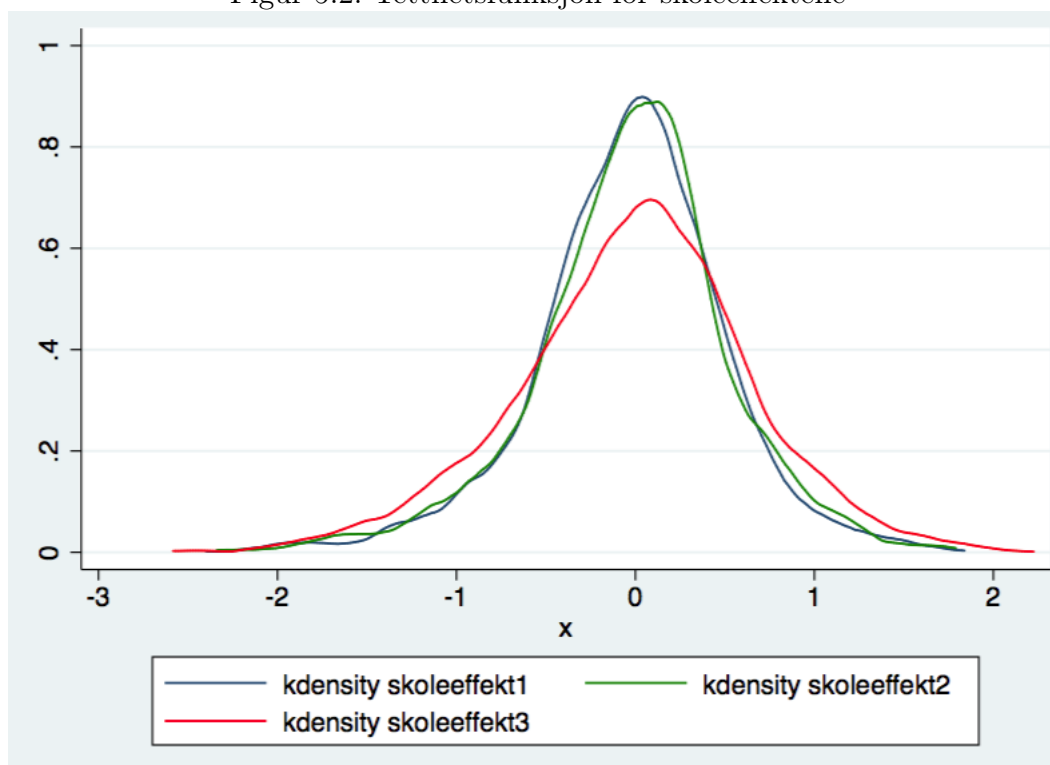
Figur 5.2 viser tetthetsfunksjoner for skoleeffektene for de ulike søskenparitetene, og fordelingen for søskenparitetene er forskjellige. Den grafiske fordelingen for tredjefødt er "lav og tykk", men for første- og andrefødte er fordelingen "høy og tynn". Sagt med andre ord er det større spredning på skoleeffektene ved tredjefødte, mens de er mer samlet for første- og andrefødte. Tetthetsfunksjonen for første- og andrefødte minner om tetthetsfunksjonen i

---

<sup>6</sup>Se ligning 2.1

<sup>7</sup>Se tabell 5.7

Figur 5.2: Tetthetsfunksjon for skoleeffektene



figur 5.1 der vi har med skoleeffekter for alle elevene og registrerer liten variasjon.

### Uteliggere

Av tabell 5.7 observerer vi differansen mellom max- og min-verdier for hver søskenparitet.

Gjengir estimatene i tabell 5.8:

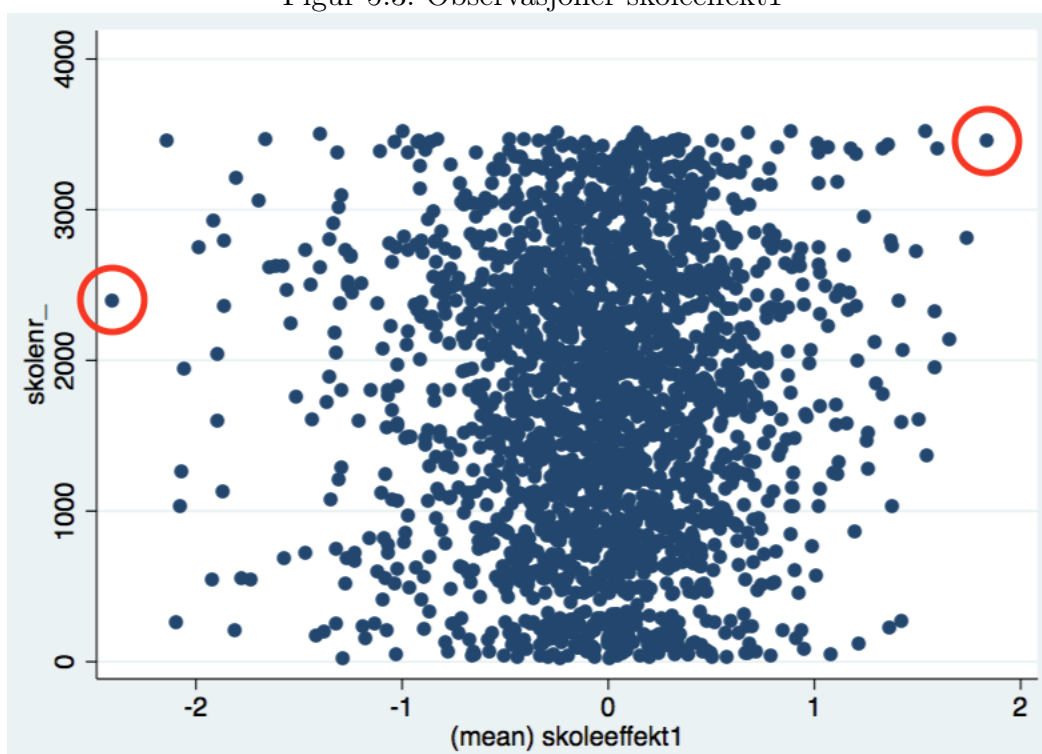
Tabell 5.8: Differanse skoleeffekter

Skoleeffekt	1	2	3
<b>Min</b>	-2.398	-2.340	-2.583
<b>Max</b>	1.838	1.790	2.225
<b>Differanse</b>	4.236	4.130	4.808

Som forventet er differansen for tredjefødte størst da vi her opplever størst variasjon. Med høy verdi på ytterpunktene har man elever som i stor grad påvirkes av skolekvalitet. Høyt

positivt tall betyr elever som gjør det godt som følge av skolekvaliteten, mens negativt det motsatte. Differansen for førstefødte er noe større enn for andrefødte, noe som ikke er som forventet når vi ser på standardavviket. Det kan dermed tyde på at vi har uteliggere<sup>8</sup> blant førstefødte. Da er det enkelte observasjoner som avviker sterkt (ekstremverdier) fra de andre observasjonene. Vi undersøker om vi har uteliggere for førstefødte ved å se på alle 2210 observasjonene for *Skoleeffekt1*. Viser observasjonene i figur 5.3.

Figur 5.3: Observasjoner skoleeffekt1



Ser av figuren at vi har uteliggere for skoleeffekten til førstefødt. De to uteliggerne har vi ringet rundt. Spesielt observasjonen på venstresiden avviker grovt fra de andre, men også en observasjon på høyresiden utmerker seg. Det er dermed en skole som har ekstremt negativ effekt på resultatene, og en som har veldig positiv effekt. Vi ønsker nå å fjerne disse to observasjonene og ser om det påvirker estimatene for *Skoleeffekt1*. Tabell 5.9 viser forskjellen på våre estimater ved å fjerne de to nevnte ekstremverdiene fra figuren.

*Skoleeffekt1* er skoleeffekten for førstefødte fra tabell 5.7, mens *Skoleeffekt1\** er den ”nye” skoleeffekten for førstefødte der vi har eliminert uteliggere. Observasjonene vi har eliminert

<sup>8</sup>På engelsk: outliers

Tabell 5.9: Forskjell skoleeffekt1

Variabel	Antall obs.	Mean	Std. Dev	Min	Max
<b>Skoleeffekt1</b>	2210	-0.0295028	0.5326607	-2.398161	1.838276
<b>Skoleeffekt1*</b>	2208	-0.0292759	0.5290195	-2.134408	1.741833
<b>Differanse</b>	2	0.0002269	0.0036412	-0.263753	-0.096443

er -2.398 og 1.838<sup>9</sup>. Variasjonen går marginalt ned som et resultat av eliminering av ytterpunktene. Da absoluttverdien av den negative observasjonen er større enn den positive,  $|-2.398| > |1.838|$ , endrer gjennomsnittet seg i positiv retning.

Den store forskjellen ser vi på max- og min-verdiene. Minimumsverdien går fra -2,398 til -2,134 og maksimumsverdien fra 1,838 til 1,742. Differansen for skoleeffekten til førstefødte har dermed endret seg fra 4,236 til 3,876, noe som tilsvarer en differanse på hele 0,36. I tabell 5.10 inkluderer vi nå de "nye" skoleeffektene for førstefødte.

Tabell 5.10: Differanse skoleeffekter u/uteliggere

Skoleeffekt	1*	2	3
<b>Min</b>	-2.134	-2.340	-2.583
<b>Max</b>	1.742	1.790	2.225
<b>Differanse</b>	3.876	4.130	4.808

Etter at vi har korrigert for uteliggere for førstefødte er differansen stigende med søskenpariteten.

Selv om differansen endrer seg utgjør ikke uteliggerne noen stor forskjell på estimatetene på skoleeffekten til førstefødte. Videre bruker vi derfor variabelen *Skoleeffekt1*<sup>10</sup> da denne variabelen har med alle observasjonene.

<sup>9</sup>Se tabell 5.8

<sup>10</sup>Skoleeffekt med uteliggere

### 5.3.1 Korrelasjonsanalyse

Nå anvendes en egen regresjon for å se om de forskjellige skolene påvirker første-, andre- og tredjefødte likt. For å finne ut om en skole som er bra for førstefødte også er bra for andre- og tredjefødte lager vi en korrelasjonsmatrise for skoleeffektene for første-, andre- og tredjefødte.

Tabell 5.11: Korrelasjonsmatrise

	Skoleeffekt1	Skoleeffekt2	Skoleeffekt3
Skoleeffekt1	1.0000		
Skoleeffekt2	0.2520	1.0000	
Skoleeffekt3	0.2312	0.2325	1.0000

Korrelasjonsmatrisen i tabell 5.11 viser oss korrelasjonskoeffisientene mellom de forskjellige skoleeffektene. Korrelasjonskoeffisienten vil variere mellom -1 og 1 og gir oss en indikasjon på hvor mye skoleeffektene korrelerer med hverandre. Hvis korrelasjonskoeffisienten er -1 eller 1 har vi perfekt korrelasjon og er korrelasjonskoeffisienten 0 har vi ingen korrelasjon. Korrelasjonen er positiv mellom alle effektene, noe som forteller oss at en økning i en av skoleeffektene også vil føre til en økning i de to resterende. Vi legger merke til at korrelasjonen mellom *Skoleeffekt1* og -3 er litt mindre enn *Skoleeffekt1* og -2, men er forholdsvis lik. Ut i fra tallene vi fikk i variansanalysen<sup>11</sup> kunne man tenke seg at korrelasjonene mellom *Skoleeffekt1* og -2 skulle vært relativt høyere enn korrelasjonen mellom *Skoleeffekt1* og -3. Vi kan alternativt tenke oss at en økning i skoleeffekt hos førstefødte fører til omtrent lik økning hos andre- og tredjefødte. For å få et bedre bilde av korrelasjonen mellom skoleeffektene lager vi en ligning med *Skoleeffekt1* som utfallsvariabel og ser hvordan *Skoleeffekt2* og -3 påvirker *Skoleeffekt1*. Anvender MKM på ligning (4.8) som ble presentert i kapittel 4:

$$Skoleeffekt_1 = \alpha_0 + \alpha_1 Skoleeffekt_t + \epsilon_t, t = 2, 3 \quad (4.8)$$

Da korrelasjonsmatrisen viser korrelasjonen mellom skoleeffektene, viser estimatene fra tabell 5.12 stigningstallet til regresjonen på ligning (4.8), der vi har *Skoleeffekt1* som avhengig

<sup>11</sup>Se tabell 5.7

Tabell 5.12: Skoleeffekt1

Variabel	1	2
<b>Skoleeffekt2</b>	<b>0.2341938</b>	
<i>Std. Err.</i>	<i>0.0205721</i>	
<b>Skoleeffekt3</b>		<b>0.1772103</b>
<i>Std. Err.</i>		<i>0.0171833</i>
<b>Konstant</b>	<b>-0.0323823</b>	<b>-0.0185801</b>
<i>Std. Err.</i>	<i>0.0107692</i>	<i>0.0107841</i>
$R^2$	<b>0.0588</b>	<b>0.0515</b>
<b>Antall observasjoner</b>	<b>2076</b>	<b>1959</b>

variabel. Øker vi for eksempel *Skoleeffekt2* med én, vil skoleeffekten for førstefødte øke i gjennomsnitt med 0,234. En økning i *Skoleeffekt2* gir større effekt på førstefødte enn en økning i *Skoleeffekt3*, men vi merker oss at konstantleddet til *Skoleeffekt2* er noe lavere enn konstantleddet til *Skoleeffekt3*. For gitt lav *Skoleeffekt1*, som vil si en skole som påvirker førstefødte lite, vil *Skoleeffekt3* være mindre negativ i forhold til *Skoleeffekt2*. Men når gitt nivå for *Skoleeffekt1* blir høyt, vil *Skoleeffekt2* overgå *Skoleeffekt3*

## 5.4 Veien videre

Analysen vår stopper her. Vi har nå estimater som viser at tredjefødte påvirkes i større grad av skolekvaliteten enn for første- og andrefødte. Et naturlig steg videre i analysen vil være å undersøke hva som driver skoleeffekten. Dette kan gjøres ved MKM med skoleeffekt som utfallsvariabel og skoleforhold som uavhengige variabler. Det kan tenkes at variabler som lærertetthet, ressursbruk og skolestørrelse er essensielle her. Funnene her ville ha kastet nytt lys på estimatene som sier at variasjonen i skoleeffekt er høyere for tredjefødte enn for førstefødte.

# Kapittel 6

## Avslutning

I denne oppgaven har vi undersøkt skolens påvirkning for elever av ulik søskenparitet. Først, sammenhengen mellom søskenpariteten og resultater på skolen, for å bekrefte at vi har et datasett hvor søskenrekkefølgeeffekten er tilstede. Deretter undersøkte vi hvordan de ulike skoleeffektene varierte for hver søskenparitet, for å kunne undersøke vår hovedhypotese om at skolekvaliteten er viktigere for elever med høy søskenparitet. Til slutt undersøkte vi korrelasjonen mellom skoleeffektene.

Vi finner en signifikant negativ effekt forbundet med søskenparitet i datasettet, selv når vi kontrollerer for familiestørrelse. Noe overraskende finner vi at familiestørrelsen kun er signifikant for tredjefødte.

Hovedresultatet i oppgaven er at skoleeffekten varierer mer for tredjefødte sammenlignet med første- og andrefødte. Vi konkluderer med at skolen har større påvirkning på elever med høyere søskenparitet. Resultatet stemmer overens med utvidelsen av modellen til Hotz & Pantano (2015) som vi utledet i kapittel 2. Vår utvidelse av modellen var at senerefødtes elevprestasjoner ville avhenge mer av skolekvaliteten. Ved å undersøke standardavvikene til skoleeffektene til første-, andre- og tredjefødte fant vi at det ikke var stor forskjell mellom første- og andrefødte, selv når vi fjernet uteliggere i dataen. Derimot fra andre- til tredjefødte gjorde standardavviket et nevneverdig hopp og ble større. Dette tolkes som at skolen er viktigere for tredjefødte i forhold til første- og andrefødte. En forklaring på dette kan ifølge



Hotz og Puntao være at foreldrene er strenge med førstefødte, men i avtagende grad utover i søskenflokket. Utvidelsen vår tilsier at skolen derfor overtar en større del av disiplineringen ovenfor senerefødte, og som følge av dette er skolekvaliteten viktigere for disse elevene.

Avslutningsvis ble korrelasjonen mellom de ulike skoleeffektene undersøkt. Her kom det frem at skoleeffektene korrelerer positivt og i nokså lik grad mellom alle de tre søskenparitetene. Positiv korrelasjon betyr at en god skole for en av søskenparitetene også er en god skole for de to andre.

Analysen ble gjort på et tverrsnitt så muligheten til å se på "within"-variasjonen i de enkelte familiene var ikke tilstede. Dette kan være en svakhet i analysen, da det er problematisk å sammenligne elever med sprikende sosioøkonomisk bakgrunn.

# Bibliografi

- Becker, G. S. (1960), An economic analysis of fertility, *in* 'Demographic and economic change in developed countries', Columbia University Press, pp. 209–240.
- Behrman, J. R. (1988), 'Nutrition, health, birth order and seasonality: Intrahousehold allocation among children in rural india', *Journal of Development Economics* **28**(1), 43–62.
- Belmont, L. & Marolla, F. A. (1973), 'Birth order, family size and intelligence', *Science* **182**, 1096–1101.
- Bjerkedal, T., Kristensen, P., Skjeret, G. A. & Brevik, J. I. (2007), 'Intelligence test scores and birth order among young norwegian men (conscripts) analyzed within and between families', *Intelligence* **35**(5), 503–514.
- Black, S. E., Devereux, P. J. & Salvanes, K. G. (2005), 'The more the merrier? the effect of family size and birth order on children's education', *The Quarterly Journal of Economics* pp. 669–700.
- Blake, J. (1981), 'Family size and the quality of children', *Demography* **18**(4), 421–442.
- Bonesrønning, H. & Massih, S. S. (2011), 'Birth order effects on young students' academic achievement', *Journal of Socio-Economics* **40**, 824–832.
- Clarke, E. L. (1916), *American men of letters, their nature and nurture*, 1 edn, Columbia University.
- Downey, D. B. (2001), 'Number of siblings and intellectual development: The resource dilution explanation.', *American Psychologist* **56**(6-7), 497.

- Elstad, E. (2009), 'Schools which are named, shamed and blamed by the media: School accountability in Norway', *Educational Assessment, Evaluation and Accountability* **21**(2), 173–189.
- Galton, F. (1874), *English men of Science Their nature and nurture*, London: Macmillan & Co.
- Hanushek, E. A. (1992), 'The trade-off between child quantity and quality', *Journal of political economy* pp. 84–117.
- Hao, L., Hotz, V. J. & Jin, G. Z. (2008), 'Games that parents and adolescents play: Risky behavior, parental reputation and strategic transfers', *Econ J* **118**, 515–555.
- Hotz, J. V. & Pantano, J. (2015), 'Strategic parenting, birth order and school performance', *Journal of Population Economics* .
- Kantarevic, J. & Mechoulan, S. (2006), 'Birth order, educational attainment, and earnings an investigation using the psid', *Journal of human resources* **41**(4), 755–777.
- Kristensen, P. & Bjerkedal, T. (2007), 'Explaining the relation between birth order and intelligence', *Science* **316**(5832), 1717–1717.
- Leman, K. (2009), *The birth order book: Why you are the way you are*, Revell.
- Leuven, E., Oosterbeek, H. & Rønning, M. (2008), 'Quasi-experimental estimates of the effect of class size on achievement in Norway\*', *The Scandinavian Journal of Economics* **110**(4), 663–693.
- McLanahan, S. (2004), 'Diverging destinies: How children are faring under the second demographic transition', *Demography* **41**(4), 607–627.
- Powell, B. & Steelman, L. C. (1990), 'Beyond sibship size: Sibling density, sex composition, and educational outcomes', *Social Forces* **69**(1), 181–206.
- Price, J. (2008), 'Parent-child quality time: does birth order matter?', *Journal of Human Resources* **43**(1), 240–265.

- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), 'Teachers, schools, and academic achievement', *Econometrica* **73**(2), 417–458.
- Rognan, A. & Barrabés, N. (2001), Nus2000, Dokumentasjonsrapport, Statistisk sentralbyrå.
- Schooler, C. (1972), 'Birth order effects: Not here, not now!', *Psychological Bulletin* **78**, 161–175.
- Seland, I., Vibe, N. & Hovdhaugen, E. (2013), 'Evaluering av nasjonale prøver som system', *NIFU-rapport 4*.
- Stinebrickner, R. & Stinebrickner, T. (2008), 'The causal effect of studying on academic performance', *BE J Econ Anal Poli* **8**.
- Strøm, B. (2004), 'Student achievement and birthday effects', *Unpublished manuscript, Norwegian University of Science and Technology*.
- Wooldridge, J. (2012), *Introductory econometrics: A modern approach*, Cengage Learning.
- Zajonc, R. B. & Markus, G. B. (1975), 'Birth order and intellectual development', *Psychological Review* **82**, 74–88.