

Gir økt skolestørrelse bedre elevresultater?
En empirisk undersøkelse av PISA i de
nordiske landene

Nicolai Ramstad

Mai 2014

Institutt for samfunnsøkonomi
Norges teknisk-naturvitenskapelige universitet

Forord

Denne oppgaven markerer avslutningen på et femårig studium i samfunnsøkonomi ved NTNU. Prosessen med å skrive oppgaven har vært veldig lærerrik, og gitt meg et innblikk i gjennomføringen og utfordringer ved samfunnsøkonomiske analyser.

Jeg ønsker å rette en stor takk til min veileder, Torberg Falch. Din veiledning har hele veien vært grundig og svært hjelpsom, og har definitivt styrket min oppgave.

Ellers ønsker jeg å takke Kristoffer for god hjelp og nyttige kommentarer. Eventuelle feil og mangler kan ikke klandres andre enn meg selv.

Nicolai Ramstad, mai 2014.

Innhold

1	Innledning	1
2	Teoretisk bakgrunn	4
2.1	Skalaavkastning	4
2.1.1	Skalafordeler	4
2.1.2	Skalaulemper	5
2.2	Produktfunksjon for utdanning	5
2.2.1	Den samtidige spesifikasjonen	6
2.2.2	Value-added spesifikasjonen	7
2.2.3	Den kumulative spesifikasjonen	8
2.2.4	Oppsummering	8
3	Litteraturoversikt	9
3.1	Skolestørrelse og kostnader	9
3.1.1	Studier på skalaavkastning	9
3.1.2	Studier på faglig tilbud	11
3.2	Skolestørrelse og elevutfall	12
3.2.1	Tidligere studier	13
3.2.2	Nyere studier	14
3.2.3	Oppsummering	15
4	Estimeringsstrategi og økonometriske utfordringer	17
4.1	Modellen	17
4.1.1	Utvalgs- og restleddsforutsetninger	17
4.2	Brudd på eksogenitetsforutsetningen	18
4.2.1	Utelatte variable	18
4.2.2	Målefeil	20
4.2.3	Seleksjon	20
4.2.4	Kontrollvariabler og proxy-variabler	21
4.2.5	Instrumentvariabel-metoden	21
4.3	Effisiens	23
4.3.1	Heteroskedastiske restledd	23
4.3.2	Korrelerte restledd	24
4.3.3	Cluster robuste standardfeil	24
5	Beskrivelse av data	25
5.1	PISA-undersøkelsen	25
5.2	Elevutfall	26
5.3	Skolestørrelse	27

5.4	Kontrollvariabler	29
5.4.1	Elev- og familiekarakteristikk	30
5.4.2	Skolekarakteristikk	31
5.4.3	Andre kontrollvariabler	31
5.5	Manglende data	32
6	Resultater	33
6.1	Minste kvadraters metode	33
6.1.1	Matematikk	33
6.1.2	Lesing og naturfag	34
6.1.3	Effekt av kontrollvariabler	34
6.1.4	Ikke-lineær funksjon	36
6.1.5	Skjeve estimater?	36
6.2	Instrumentvariabel-estimering	38
6.2.1	Matematikk	38
6.2.2	Lesing og naturfag	39
6.2.3	Oppsummering	40
6.3	Heterogene effekter	41
6.3.1	Enkeltland	41
6.3.2	Innvandringsbakgrunn	42
6.3.3	Sosioøkonomisk bakgrunn	43
6.3.4	Kjønn	44
6.3.5	Oppsummering	45
7	Oppsummering og diskusjon	47
A	Appendiks	52
A.1	Deskriptiv statistikk	52
A.2	Regresjoner	56
A.3	Figurer	67

Figurer

1	Eksempler på skalafordeler og skalaulemper	4
2	Utvikling i matematikkresultater i Norden	27
3	Utvikling i leseresultater i Norden	27
4	Utvikling i naturfagsresultater i Norden	28
5	Histogram over skolestørrelse	28
6	Spredningsplott med skolestørrelse og matematikkresultater	29
7	Spredningsplott med skolestørrelse og matematikkresultater hvor skolestørrelse over 1000 er fjernet	67
8	Sammenheng mellom skolestørrelse og matematikkresultater med kvadratisk ledd for skolestørrelse	67

Tabeller

1	MKM-estimering med matematikkresultater som avhengig variabel	34
2	MKM-estimering med matematikkresultater som avhengig variabel, inkludert ledd for kvadrert skolestørrelse	37
3	IV-estimering med matematikkresultater som avhengig variabel	39
4	IV-estimering med matematikkresultater som avhengig variabel, enkeltland	42
5	IV-estimering med matematikkresultater som avhengig variabel, innvandringsbakgrunn	43
6	IV-estimering med matematikkresultater som avhengig variabel, sosioøkonomisk bakgrunn (SES)	44
7	IV-estimering med matematikkresultater som avhengig variabel, kjønn	45
A1	Deskriptiv statistikk	52
A2	Deskriptiv statistikk, skolestørrelse	55
A3	MKM-estimering med leseresultater som avhengig variabel	56
A4	MKM-estimering med naturfagsresultater som avhengig variabel	56
A5	MKM-estimering med alle kontrollvariabler inkludert. Avhengige variabler er testresultater i de tre fagene.	57
A6	IV-estimering med leseresultater som avhengig variabel	61
A7	IV-estimering med naturfagsresultater som avhengig variabel	62
A8	IV-estimering med leseresultater som avhengig variabel, enkeltland	63
A9	IV på naturfagsresultater, enkeltland	63
A10	IV-estimering med leseresultater som avhengig variabel, innvandringsbakgrunn	64
A11	IV-estimering med naturfagsresultater som avhengig variabel, innvandringsbakgrunn	64

A12	IV-estimering med leseresultater som avhengig variabel, sosioøkonomisk bakgrunn (SES)	65
A13	IV-estimering med naturfagsresultater som avhengig variabel, sosioøkonomisk bakgrunn (SES)	65
A14	IV-estimering med leseresultater som avhengig variabel, kjønn	66
A15	IV-estimering med naturfagsresultater som avhengig variabel, kjønn	66

1 Innledning

I mange år har utformingen av utdanningssystemet vært i fokus for hvordan man på best mulig måte kan sikre tilstrekkelig tilbud av godt kvalifisert arbeidskraft. Innen utdanningsøkonomi er det mange teorier, men ingen konsensus om hvordan man kan optimalisere utformingen av skolesektoren for å sikre en effektiv læringshverdag for elevene. I forbindelse med offentliggjøringen av resultatene fra PISA-undersøkelsen i 2012, har debatten igjen blusset opp om hvordan utdanningssystemet best kan utformes for å bedre elevresultatene. Av forslag som har blitt foreslått og innført i visse land, er kanskje redusert klassestørrelse, accountability-systemer og fritt skolevalg de mest fremtredende. Økt ressursbruk har i mange tilfeller vist seg ikke å være et tilstrekkelig virkemiddel for bedret skolekvalitet, og de landene som bruker mest ressurser på utdanning er ikke nødvendigvis de som gjør det best på internasjonale tester. Et eksempel er Norge, som ligger i toppen av OECD landene på ressursbruk på utdanning (OECD, 2013), men som ikke har prestert mer enn på OECD gjennomsnittet, eller noe under, på PISA-undersøkelsene på 2000-tallet.

I skoledebatten er mye oppmerksomhet blitt rettet mot klassestørrelse, men et annet sentralt spørsmål er i hvilken grad skolestørrelse påvirker elevprestasjoner. I USA, som er det landet hvor det er blitt forsket mest på skolestørrelse, har argumentene for å redusere størrelsen på skoler blitt flere siden 1990-tallet. Mye av den empiriske forskningen på området finner at elever ved små skoler gjør det relativt bedre enn elever ved store skoler, blant annet på områder som frafallsrater og sannsynligheten for å fullføre utdanningen (Fowler Jr. og Walberg, 1991; Kuziemko, 2004; Barrow et al., 2013). Basert på denne forskningen har mange av de største byene i USA gjennomført reformer for å redusere skolestørrelsen, inkludert New York og Chicago. Nyere forskning som ser på effektene av disse reformene, blant annet Schwartz et al. (2013), finner at de skolene som presterer best på en rekke utfallsmål er små nyopprettede skoler. Schwartz et al. (2013) finner også at små skoler som eksisterte før reformen ble innført ikke presterer bedre enn store skoler. Det er svært interessant å finne ut hva som ligger bak de gode resultatene til de små nyopprettede skolene. Er det størrelsen som teller, eller er det andre strukturelle forhold ved disse skolene som ligger bak de gode resultatene?

Forkjempere for små skoler argumenterer for en økt følelse av tilhørighet for elevene, mindre mobbing og vold fordi lærere og elever kjenner hverandre bedre, samt høyere forventninger til elevene (Meier, 1996; Pittman og Haugwout, 1987). Det er også blitt foreslått at det på små skoler foregår en mer effektiv selvevaluering av lærere, bedre samhandling mellom lærere, og større foreldreinvolvering (Cotton, 1996; Jewell, 1989). På den andre siden av debatten argumenteres det for kostnadsbesparelser og et bredere faglig tilbud ved økt skolestørrelse (Conant, 1959). Lavere kostnader per elev, eller stordrifts-

fordeler, kan oppstå for større skoler blant annet ved mer effektiv utnyttelse av fysiske anlegg, spredning av uatskillelige kostnader, og mer spesialisering av arbeidskraften (Lee og Smith, 1997; Buzacott, 1982). Det argumenteres for at disse kostnadsbesparelsene kan benyttes til å tilby et bredere faglig tilbud.

I Norge har det vært en del debatt om hva som er optimal skolestørrelse, særlig i forbindelse med byggingen av nye skoler (for eksempel i Kongsvinger), og også i forbindelse med nedleggelse og sammenslåinger av skoler. I perioden 2005 til 2013 ble det lagt ned over 350 offentlige grunnskoler (Dagbladet, 2013). Dette var i hovedsak skoler med lavt elevtall, og halvparten av disse skolene hadde under 34 elever. Med manglende forskning og empiri på området i Norge blir det derfor naturlig å stille spørsmål ved hva slags grunnlag beslutningstagerne hadde for å ta avgjørelsen om nedleggelse av disse skolene.

Sollien (2008) tar opp spørsmålet om skolestørrelse i norsk sammenheng i sin rapport, hvor hun oppsummerer den eksisterende litteraturen på området, og belyser den manglende forskningen på dette i Norge. Hun påpeker viktigheten av slik forskning, fordi studier fra USA ikke nødvendigvis er generaliserbare til Norge og Norden grunnet store forskjeller i utdanningssystemene. De amerikanske skolene er ofte store institusjoner med langt flere elever enn hva som er vanlig i Norden. De skiller seg også fra de nordiske skolene ved at de i langt større grad inkluderer idrett og andre aktiviteter i tillegg til faglig læring, noe som gir en forskjell i hvordan elevene opplever skoletilhørighet. Det er også en større andel elever med minoritetsbakgrunn i amerikanske skoler, samt en større variasjon i sosioøkonomisk bakgrunn. Derfor vil ikke nødvendigvis konklusjonen om at redusert skolestørrelse gir bedre elevprestasjoner gjelde for det nordiske skolesystemet. For eksempel er det grunn til å tro at det ikke gir den samme effekten å redusere (øke) størrelsen på en skole med 1000 elever, som å redusere (øke) størrelsen på en skole med 100 elever.

I denne oppgaven vil jeg se på hvordan skolestørrelse påvirker elevresultater på internasjonale tester i Norden, eksklusiv Island. Jeg vil bruke elev- og skoledata fra tre år med PISA-undersøkelser. Siden det ikke er de samme elevene og skolene som undersøkes i de tre undersøkelsene, er det ikke et vanlig paneldatasett jeg benytter, men jeg bruker data fra flere år for å øke antall observasjoner og for å kunne se på forskjeller over tid. Lite forskning er blitt gjort på skolestørrelse i Norden, og det er derfor spennende å se i hvilken grad de amerikanske konklusjonene er generaliserbare til de nordiske landene. Fordi gjennomsnittlig skolestørrelse i Norden er forholdsvis lav, er min hypotese at en økning i skolestørrelse vil gi en mer effektiv produksjon av skoletjenester og bedre elevresultater, blant annet på grunn av stordriftsfordeler. Helt konkret er problemstillingen jeg ønsker å undersøke:

Gir økt skolestørrelse i de nordiske landene bedre resultater for elevene?

Det er av interesse å undersøke denne problemstillingen, fordi dersom skoler er for små,

kan en legge til rette for at resultater forbedres og kostnader reduseres ved å slå sammen skoler eller bygge større skoler. Det kan også være et argument for eller mot kommunesammenslåinger dersom dette medfører skolesammenslåinger.

PISA-undersøkelsen kartlegger prestasjonene til 15-åringer, som betyr at jeg vil se på sammenhengen mellom skolestørrelse og elevprestasjoner på ungdomskolenivå. Dette er viktig å presisere, fordi det ikke er en selvfølge at de samme skolestørrelseeffektene vil gjelde for barneskoler, ungdomsskoler og videregående skoler. I litteraturen er det blitt funnet at store skoler kan være positivt for eldre elever, mens mindre skoler er bedre for yngre elever (Friedkin og Necochea, 1988).

Utgangspunktet for min analyse vil være minste kvadraters metode (MKM) av en skoleproduktfunksjon. Fordi denne metoden kan by problemer på grunn av utelatte variable og selvseleksjon, vil jeg videre benytte meg av instrumentvariabel-metoden (IV) i et forsøk på å håndtere disse problemene. Metoden jeg bruker er motivert av tidligere studier (Barrow et al., 2013; Schwartz et al., 2013), hvor det benyttes instrumentvariabler som baserer seg på avstand til nærmeste skoler.

Opgaven er bygd opp slik at jeg i kapittel 2 går gjennom aktuell teori for oppgaven, mens jeg i kapittel 3 vil gi en oversikt over eksisterende litteratur på skolestørrelse. I kapittel 4 ser jeg på metode og økonometriske utfordringer, og i kapittel 5 gir jeg en beskrivelse av datasettet jeg benytter. I kapittel 6 presenterer jeg resultatene fra analysen, før jeg til slutt konkluderer i kapittel 7.

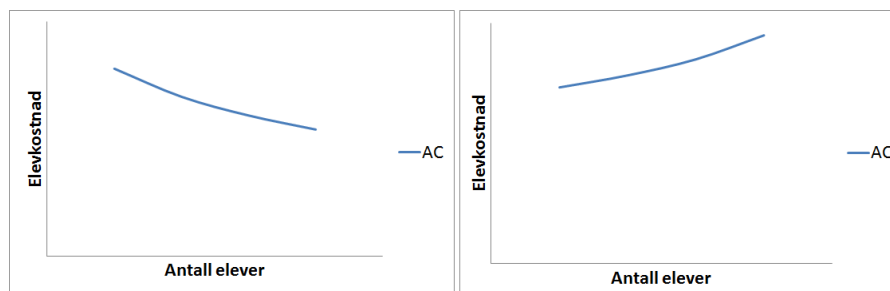
2 Teoretisk bakgrunn

Jeg vil i dette kapitlet gå gjennom litt av teorien som er relevant for oppgaven. I neste delkapittel vil jeg presentere litt av teorien rundt skalaavkastning, siden dette er brukt som et av hovedargumentene blant talsmenn for økt skolestørrelse. Videre vil jeg beskrive utformingen av produktfunksjoner for utdanning, hvor jeg vil belyse fordeler og svakheter ved forskjellige tilnærminger, inkludert den som benyttes i min analyse.

2.1 Skalaavkastning

Skalaavkastning er et begrep innen produksjonsteorien som beskriver hva som skjer med en bedrifts gjennomsnittskostnader når produksjonen øker. Figur 1 viser eksempler på skalafordeler og skalaulemper i produksjonen, og jeg vil i det videre forklare disse konseptene grundigere. Teorien i dette delkapitlet er hentet fra læreboka til Begg et al. (2008).

Figur 1: Eksempler på skalafordeler og skalaulemper



2.1.1 Skalafordeler

I en skolesammenheng kan skalafordeler oppstå av to grunner. For det første på grunn av udelelige kostnader i produksjonsprosessen. En skole vil ha en del faste kostnader uavhengig av elevtallet, for eksempel de mest nødvendige ansatte eller skolebygninger. Hvis en starter fra et lavt elevtall, vil ikke disse kostnadene øke når en gradvis øker elevtallet, og kostnadene vil spres utover flere elever. Gjennomsnittskostnadene reduseres som vist til venstre i figur 1. Etter hvert som skolene når høyere elevtall, må de kanskje ansette flere, og skalafordelene forsvinner.

Den andre grunnen til skalafordeler er spesialisering. Etter hvert som en bedrift ansetter flere personer, kan hver enkelt ansatt få mer spesialiserte oppgaver. På en liten skole med veldig få lærere, må kanskje hver lærer undervise i alle fag uavhengig av hva de har sterkest kompetanse i. Etter som skolestørrelsen øker, øker antallet lærere, og hver enkelt lærer kan undervise i de fagene de kan best.

2.1.2 Skalaulemper

Etter hvert som elevtallet på en skole øker, kan det oppstå skalaulemper og en situasjon med økende gjennomsnittskostnader, som vist til høyre i figur 1. Som med et hvilket som helst firma blir styringen vanskeligere med økt størrelse, da dette kan gi økt byråkrati og koordineringsproblemer.

En del av problemene med store skoler som er nevnt i litteraturen vil ganske sikkert også føre med seg økte kostnader. Det er blitt foreslått at elever på store skoler i mindre grad føler tilhørighet, samt at det er mer vold på store skoler fordi lærere og elever kjenner hverandre dårligere. Mindre foreldreinvolvering, dårligere samhandling mellom lærere, og mindre effektiv selvevaluering av lærere (Cotton, 1996; Fowler Jr. og Walberg, 1991; Jewell, 1989) kan bidra til å øke kostnadene for store skoler relativt til små, eller i det minste gi dårligere resultater til samme kostnadsnivå.

Etter hvert som elevtallet øker, avhenger formen på gjennomsnittskostnadskurven av to ting: hvor lenge skalafordelene varer og når skalaulempene inntreffer.

2.2 Produktfunksjon for utdanning

På samme måte som bedrifters teknologi kan presenteres ved en produktfunksjon, hvor man ser på sammenhengen mellom innsatsfaktorer og produksjonsnivå, kan elevers prestasjoner presenteres som en funksjon av en rekke innsatsfaktorer som påvirker prestasjonene. En vanlig produktfunksjon for utdanning relaterer innsatsfaktorer som påvirker elevers læring, som for eksempel skole, familie, medelever og miljø, med diverse utfallsvariabler, som for eksempel prøveresultater, uteksamineringsrater, suksess i arbeidslivet eller sannsynligheten for å ta videre utdanning.

Et viktig bidrag til litteraturen på produktfunksjoner for utdanning er artikkelen til Todd og Wolpin (2003), og teorien som presenteres i dette avsnittet er hentet fra deres artikkel. Kort oppsummert, beskriver de metoder for å modellere produktfunksjoner for kognitive ferdigheter ved å ta hensyn til at barns utvikling er en kumulativ prosess som avhenger av historiske innsatsfaktorer fra foreldre og skolen, og av medfødte evner. Analysen tar utgangspunkt i en helt generell produktfunksjon for utdanning

$$T_{ija} = T_a [F_{ij}(a), S_{ij}(a), \mu_{ij0}, \epsilon_{ija}] . \quad (1)$$

Her er T_{ija} et mål på prestasjoner for elev i fra husholdning j ved alder a . $F_{ij}(a)$ er en vektor av foreldres innsatsfaktorer opp til alder a , mens $S_{ij}(a)$ er en vektor av skolens innsatsfaktorer opp til alder a . μ_{ij0} er et mål på elevens medfødte evnenivå, eller men-

tale kapasitet. ϵ_{ija} er restleddet, og ved å inkludere dette tas det høyde for målefeil i testresultatene.

Modellen i (1) kan ses på som en ideell produktfunksjon for utdanning, men det er dessverre en del problemer som gjør at det er vanskelig å estimere den. For det første er det vanskelig å finne et mål på medfødt mental kapasitet. For det andre vil estimering av modellen kreve enorme mengder data, og de færreste datasett inneholder komplett historisk informasjon på alle relevante innsatsfaktorer. I stedet blir man nødt til å finne alternative spesifikasjoner som gjør at (1) blir estimert på en konsistent måte, til tross for at en del informasjon mangler.

2.2.1 Den samtidige spesifikasjonen

Den samtidige spesifikasjonen relaterer elevprestasjoner og mål på skole- og familieinnsatsfaktorer ved bruk av samtidige data, det vil si at historiske data på innsatsfaktorer ikke inkluderes. Spesifikasjonen kan brukes dersom vi antar at bare samtidige innsatsfaktorer påvirker prestasjoner i dag, at innsatsfaktorene ikke endres over tid, og at innsatsfaktorene ikke er korrelert med medfødt evnenivå.

Den samtidige spesifikasjonen kan skrives som

$$T_{ija} = T_a(F_{ija}, S_{ija}) + \epsilon_{ija}^*. \quad (2)$$

I (2) er variablene de samme som i (1), bortsett fra restleddet. Restleddet inkluderer nå også utelatte variabler, som her vil være tidligere brukte innsatsfaktorer og medfødt mental kapasitet.

Bruken av den samtidige spesifikasjonen stiller strenge forutsetninger for at man skal kunne oppnå konsistente estimatorer, og det er ikke veldig sannsynlig at noen av de nevnte forutsetningene over vil gjelde. For eksempel er det mye som tyder på at det eksisterer en link mellom erfaringer man gjør seg tidlig i barndommen og sosiale, kognitive og atferdsutfall senere, og mange av innsatsfaktorene vil ganske sikkert variere over tid og med barnets alder. Også forutsetningen om at innsatsfaktorer ikke er korrelert med medfødte egenskaper er det grunn til å sette spørsmålstegn ved.

Fordelen med denne spesifikasjonen er at den krever relativt lite data, og den benyttes som regel kun når data på historiske innsatsfaktorer og testresultater ikke foreligger.

2.2.2 Value-added spesifikasjonen

Value-added spesifikasjonen relaterer en utfallsvariabel på prestasjoner med samtidige mål på innsatsfaktorer, og et tilbakedatert mål på prestasjoner. Spesifikasjonen blir som regel tatt i bruk når data på historiske innsatsfaktorer og medfødt mental kapasitet mangler.

Helt generelt kan vi skrive value-added modellen som

$$T_{ija} = T_a \{X_{ija}, T_{a-1}[X_{ij}(a-1), \mu_{ij0}], \eta_{ija}\}. \quad (3)$$

Her er X , for enkelthets skyld, en vektor av familie- og skoleinnsatsfaktorer, og $X(a)$ historiske innsatsfaktorer opp til alder a .

Ved å anta at (3) er additiv separabel og at parameterne ikke varierer med alder, kan vi skrive value-added spesifikasjonen som

$$T_{ija} = X_{ija}\alpha + \gamma T_{ij,a-1} + \nu_{ija}. \quad (4)$$

Videre er det interessant å se på hva slags forutsetninger som må pålegges for at value-added funksjonen skal gi konsistente estimatorer. Jeg tar utgangspunkt i en utskrevet versjon av den generelle produktfunksjonen (1)

$$T_{ija} = X_{ija}\alpha_1 + X_{ija-1}\alpha_2 + \dots + X_{ij1}\alpha_a + \beta\mu_{ij0} + \epsilon_{ija}. \quad (5)$$

Denne formuleringen pålegger forutsetningen om at $T_a(\cdot)$ -funksjonen ikke avhenger av alder, i det minste ikke i tidsintervallet som brukes i implementeringen av value-added modellen. I (5) tillates effektene av innsatsfaktorene å variere avhengig av når de ble tatt i bruk. Ved å trekke fra $\gamma T_{ij,a-1}$ fra begge sider av (5) og flytte om, kan vi skrive

$$\begin{aligned} T_{ija} - \gamma T_{ij,a-1} &= X_{ija}\alpha_1 + X_{ija-1}(\alpha_2 - \gamma\alpha_1) + \dots + X_{ij1}(\alpha_a - \gamma\alpha_{a-1}) \\ &\quad + (\beta_a - \gamma\beta_{a-1})\mu_{ij0} + \epsilon_{ija} - \gamma\epsilon_{ij,a-1}. \end{aligned} \quad (6)$$

For at (6) skal kunne reduseres til (4) må vi anta at koeffisientene på innsatsfaktorene må være geometrisk avtagende med avstand i tid fra prestasjonsmålet, og at effekten av medfødte egenskaper må være geometrisk avtagende med samme rate som effektene av innsatsfaktorene, det vil si at $\beta_a = \gamma\beta_{a-1}$.

Value-added spesifikasjonen blir ofte sett på som bedre enn den samtidige spesifikasjonen

(Hanushek, 2003; Krueger, 2000), men som vist over, krever også denne at vi legger strenge forutsetninger på produksjonsteknologien. Value-added spesifikasjonen krever bare noe mer data enn den samtidige spesifikasjonen, nemlig et tilbakedatert testresultat.

2.2.3 Den kumulative spesifikasjonen

For å kunne estimere den kumulative spesifikasjonen gitt ved (1) direkte, kreves data på både historiske og samtidige familie- og skoleinnsatsfaktorer. Hvis man har data på dette, vil problemet være å finne et mål på medfødte evner. Dersom man antar at alle utelatte variabler og målefeil i testresultater er ukorrelet med inkluderte variabler, vil problemet med å estimere (1) være at valg av innsatsfaktorer kan føre til korrelasjoner mellom de observerbare innsatsfaktorene og uobserverbare evner hos barnet. I litteraturen benyttes metoder som bruker observasjoner på flere barn innen samme husholdning for å håndtere heterogenitet i medfødte evner, mens en annen metode benytter flere mål på samme barn ved forskjellige aldre for å håndtere det samme problemet. Siden jeg ikke benytter disse metodene i min analyse, og teorien rundt dette er ganske omfattende, nøyer jeg meg med kun å nevne de her.

2.2.4 Oppsummering

I dette delkapittelet har jeg beskrevet flere utforminger av produktfunksjoner for utdanning. Den samtidige spesifikasjonen krever minst data, men det stilles strenge forutsetninger for at denne skal gi konsistente estimatorer, og value-added spesifikasjonen og den kumulative spesifikasjonen er bedre alternativer dersom mer data foreligger. Jeg må allikevel nøye meg med å benytte den samtidige spesifikasjonen i min analyse på grunn av strukturen på mitt datasett. Som beskrevet over er ikke dette problemfritt, og vil by på en del økonometriske problemer. Særlig vil utelatte variabler være en stor bekymring. Dette forsøker jeg å ta høyde for ved og benytte meg av instrumentvariabel-metoden (IV). Både utelatte variabler og IV-metoden beskrives grundig i kapittel 4.

3 Litteraturoversikt

Helt siden 1950-tallet har det blitt diskutert hva som er optimal skolestørrelse. Conant (1959) foreslo i sin rapport at mange amerikanske skoler på den tiden var for små til å kunne tilby et diversifisert fagtilbud, og hvis de gjorde det, var det til alt for høye kostnader. Barker og Gump (1964) var blant de første som foreslo at skoler bør være mindre, og i en størrelsesorden hvor ingen av elevene er overflødige. Utover 1980- og 1990-tallet økte forskningen og antall publiseringer på området, med talsmenn både for og imot mindre skoler.

Litteraturen på skolestørrelse kan deles inn i to grupper; de studiene som fokuserer på kostnader, og de studiene som fokuserer på prestasjoner og elevutfall. Det er både teoretiske og empiriske studier på begge områder. Det eksisterer en rekke teorier og hypoteser som både taler for og imot økt skolestørrelse. Jeg vil i dette kapitlet oppsummere noe av den eksisterende litteraturen. Jeg vil først se på studier som baserer seg på skolestørrelse og kostnader, og deretter studier som ser på sammenhengen mellom skolestørrelse og elevprestasjoner. Mye av forskningen på skolestørrelse er gjennomført i USA, så hovedvekten av studiene jeg presenterer er amerikanske. I den grad det er mulig og relevant, vil jeg presentere studier fra andre land.

3.1 Skolestørrelse og kostnader

Talsmenn for økt skolestørrelse har særlig lent seg på to hovedargumenter for hvorfor store skoler er bedre enn små. For det første at store skoler er mer økonomisk effektive på grunn av skalafordeler, og for det andre at store skoler har muligheten til å tilby flere fag. Argumenter for og imot disse to argumentene, samt resultater fra empirisk forskning gjennomgås i det videre.

3.1.1 Studier på skalaavkastning

Oppfatningen om at store skoler har lavere kostnader per elev har grobunn i teorien om skalafordeler, som oppsummert i delkapittel 2.1. Dersom man øker alle innsatsfaktorer proporsjonalt, for eksempel kapital, arbeidskraft, læremidler og elever, kan dette i noen tilfeller gi lavere kostnader per elev, eller skalafordeler (Stiefel et al., 2000). Skalafordeler kan oppstå for store skoler av en rekke grunner, blant annet på grunn av at:

- Noen typer fysiske anlegg brukes mer effektivt ved økt størrelse, for eksempel gym-saler og kantiner (Kenny, 1982; Buzacott, 1982).

- Noen innsatsfaktorer, for eksempel administratorer, kan være uatskillelige, og kostnadene ved disse spres over flere elever (Kenny, 1982; Buzacott, 1982)
- Spesialisering og inndeling av arbeidskraft kan forekomme, for eksempel når lærere kun tilbyr undervisning i de fagene de kan best (Conant, 1959).

Studier har vist at de teoretiske argumentene for skalafordeler ikke nødvendigvis gjelder i praksis, selv ikke i en bedriftssammenheng (Gooding og Wagner, 1985). Det kan hende at økte administrative kostnader kan motvirke skalafordeler på grunn av ledelsens manglende evne til å styre store skoler (Haller, 1992; Begg et al., 2008). Akademisk og fiskal ansvarlighet er også mer sannsynlig ved mindre skoler på grunn av enklere organisering og mer synlige lærere og administrasjon (Meier, 1996)

Med argumenter både for og imot skalafordeler, er det interessant å se på hva den empiriske forskningen viser.

McGuffey og Brown (1978) ser på sammenhengen mellom skolestørrelse og utnyttelsen av fasiliteter ved skoler. De finner at store skoler utnytter skolens fasiliteter på en bedre måte, og at driftskostnader per elev er lavere ved store skoler.

Fox (1981) gjør et viktig funn i sin analyse av den eksisterende litteraturen på skolestørrelse og kostnader. Han finner at kostnadskurver kan være U-formet, noe som medfører at økt skolestørrelse forbi et optimalt nivå kan føre til økte kostnader per elev. Dette betyr at både veldig små og veldig store skoler er dyre i drift.

Stiefel et al. (2000) kritiserer mye av den eksisterende litteraturen som ser på skolestørrelse og kostnader, og nevner særlig tre store svakheter ved tidligere studier. For det første er det få studier som tar hensyn til at elevutfall og kostnader bestemmes simultant. Alt annet likt er det naturlig å anta at elevers prestasjoner er bedre når utgiftene ved skolen er høye. For det andre antas det som regel at skoler minimerer kostnadene når kostnadsfunksjoner estimeres, noe som ikke nødvendigvis er tilfellet i offentlige skoler. Til slutt kritiserer de bruken av kostnader på skoledistriktnivå i stedet for på skolenivå, som de mener er mer passende. Forbedringen i deres studie sammenlignet med tidligere studier, er at de benytter seg av et utfallsmål som kombinerer prestasjoner og kostnader, nemlig budsjett per uteksaminerte elev. De finner at små skoler er vel så kostnadseffektive som store når dette målet benyttes.

Taylor og Bradley (2000) benytter et datasett av engelske ungdomsskoler for å se på sammenhengen mellom skolestørrelse og kostnader. De bruker de ansattes arbeidstimer per elev som en proxy for kostnader per elev. De finner en signifikant negativ sammenheng mellom skolestørrelse og kostnader per elev, som indikerer at store skoler er mer kostnadseffektive enn små.

Falch et al. (2008) benytter et paneldatasett på norske skoler og kommuner over perioden 2001 til 2006 for å se på sammenhengen mellom skolestørrelse og ressurser allokert til skoler. De finner at økt skolestørrelse er forbundet med færre lærertimer per elev, og sammenhengen virker å være ikke-lineær. Siden lærerlønninger utgjør omtrent 70 prosent av totale kostnader, er dette et viktig funn, og er i tråd med resultatene til Taylor og Bradley (2000). Resultatene tyder på at store skoler er mer kostnadseffektive enn små i Norge, og at sammenslåinger av skoler kan bidra til kostnadsbesparelser i skolesektoren. De påpeker forøvrig at spørsmålet om sammenslåinger av skoler også avhenger av sammenhengen mellom skolestørrelse og elevprestasjoner, noe jeg ser nærmere på i denne oppgaven.

3.1.2 Studier på faglig tilbud

Et annet viktig argument blant talsmenn for store skoler er at det med økt størrelse følger et bredere og mer fullstendig fagtilbud (Conant, 1959; Howley, 1997).

Conant (1959) gjør en undersøkelse av videregående skoler (high schools) i en rekke stater i USA. Han finner at det er mulig for små skoler å tilby et tilstrekkelig faglig tilbud til alle elever, men at dette vil medføre svært høye kostnader. Han anbefaler skoler som uteksaminerer færre enn 100 elever å øke størrelsen.

Barker og Gump (1964) undersøker videregående skoler i USA. De finner at økt skolestørrelse ikke nødvendigvis medfører et økt faglig tilbud. Blant annet finner de at mye av det pensumet som undervises i spesialiserte fag ved store skoler, er inkludert i vanlige fag ved små skoler.

Monk (1987) bruker et datasett med skoler i New York for å se på om store skoler har større grad av ressursspesialisering og et mer helhetlig fagtilbud enn mindre skoler. Han finner at skalafordelene virker å forsvinne når elevtallet når ganske lave nivåer, og at fordelene ved økt fagtilbud er små utover dette elevtallet. Mer konkret finner han et vendepunkt når skoler når et elevtall på 400. For skoler med under 400 elever vil en økning i størrelsen føre til et bredere fagtilbud og økt spesialisering. Over dette elevtallet finner han små fordeler ved økt skolestørrelse.

Walberg og Walberg (1994) bruker også skoledata fra USA, men ser på skoledistrikter i stedet for skoler. De finner at store skoledistrikter tilbyr et bredere faglig tilbud enn små skoledistrikter. Allikevel viser det seg at det noe smalere faglige tilbudet ved små skoler kan være positivt, fordi det fører til at alle elevene tar viktige kjernefag. I tillegg påpekes det at de spesialiserte fagene ved store skoler kun vil være forbeholdt en liten del av elevene, for eksempel flinke elever eller elever med spesielle behov, og at fordelene ved disse fagene derfor er små for flertallet av elevene.

Det viser seg altså å være sprikende resultater på dette feltet i den empiriske forskningen.

Økt skolestørrelse kan ikke sies å være en garanti for et økt faglig tilbud, og dersom økt skolestørrelse fører med seg et bredere fagtilbud, betyr ikke dette at flertallet av elevene kan utnytte seg av det.

3.2 Skolestørrelse og elevutfall

Interessen i mange studier, inkludert denne, er i hvilken grad skolestørrelse påvirker elevprestasjoner. Bak skalafordelargumentet ligger forutsetningen om at de reduserte utgiftene per elev som kan oppstå ved økt skolestørrelse, vil føre til bedre prestasjoner, fordi flere ressurser kan investeres i å forbedre skolen.

Talsmenn for store skoler peker på en del psykologiske fordeler ved store skoler, ved siden av skalafordelargumentet. Smith og DeYoung (1988) nevner tre argumenter. For det første at små skoler kun tiltrekker seg elever fra et begrenset geografisk område, og at det derfor vil oppstå en mer homogen elevmasse ved disse skolene sammenlignet med større skoler. For det andre vil interaksjon med kun et lite antall lærere virke negativt på elevene ved små skoler, fordi lærerens forventninger til eleven da kan være bestemt fra første år av, noe som kan hindre utvikling og muligheten for en ny start med en ny lærer. For det tredje vil en begrenset elevmasse hindre elevens sosiale utvikling. Ved store skoler, hvor det finnes flere sosiale grupper, vil det være en større sjanse for at elevene finner en gruppe hvor de hører hjemme og føler seg komfortable i. Argumentene til Smith og DeYoung (1988) har møtt en del motstand, og et av motargumentene er at spesialisering ved store skoler vil føre til at elever må forholde seg til langt flere lærere, hvor ingen av lærerne kjenner elevene særlig godt (Strang, 1987).

I den teoretiske litteraturen er det foreslått mange grunner til hvorfor små skoler kanskje fremmer bedre elevresultater. De fleste fokuserer på det sosiale aspektet ved små skoler. Flere studier har vist at både elever og lærere har vist en større grad av tilfredshet ved små skoler, og at mindre skoler gir en økt følelse av tilhørighet (Cotton, 1996; Pittman og Haugwout, 1987; Lee et al., 1993; Barrow et al., 2013). Grunner til økt tilhørighet ved små skoler kan blant annet være at elever og lærere ved disse skolene kjenner hverandre bedre, at mindre anonymitet reduserer atferdsproblemer og vold, og at mindre skoler virker mindre skremmende på foreldre (Meier, 1996). Det er også blitt funnet en større grad av samarbeid mellom elever og lærere ved små skoler enn ved store skoler (Edington og Gardener, 1984). Andre studier har funnet at det er mindre disiplinproblemer ved små skoler (Haller, 1992), økt deltagelse i frivillige aktiviteter (Barker og Gump, 1964), og større foreldreinvolvering (Meier, 1996).

Jeg vil i det videre gå gjennom noen av studiene som har sett på sammenhengen mellom skolestørrelse og elevutfall. Først går jeg gjennom litt av litteraturen fra andre halvdel

av forrige århundre, hvor de fleste studiene baserte seg på enkle statistiske metoder som i dag blir sett på som mindre troverdige. Deretter vil jeg gå gjennom litt av den nyere forskningen på skolestørrelse og elevutfall som benytter mer avanserte metoder.

3.2.1 Tidligere studier

Fowler Jr. og Walberg (1991) bruker regresjonsanalyse på et datasett av offentlige skoler i New Jersey. De estimerer sammenhengen mellom skolestørrelse og 18 forskjellige skoleutfall, blant annet gjennomsnittresultatet på statlige tester, elevenes gjennomføringsgrad, og utvisninger. De finner gjennomgående en negativ sammenheng mellom skolestørrelse og elevutfall. Resultatene indikerer at elever ved små skoler presterer bedre langs en rekke utfall, uavhengig av sosioøkonomisk bakgrunn og alder.

Lee og Smith (1997) benytter et paneldatasett bestående av elever fra offentlige, katolske og private skoler i USA. De bruker en hierarkisk lineær modell (HLM) for å undersøke sammenhengen mellom elevenes prestasjonsvekst i matte og lesing, og skolestørrelse i amerikanske videregående skoler. Ut i fra resultatene anbefaler de en skolestørrelse på mellom 600 og 900 elever.

Page et al. (2002) oppsummerer den eksisterende litteraturen på skolestørrelse, med formål om å komme til en konklusjon angående skolestørrelse og beste utforming av små læringsmiljøer. De gjør et dypdykk i 55 artikler, som inkluderer kostnadsanalyser, kvalitative case-studier, og utfallsbaserte studier. Basert på den eksisterende litteraturen konkluderer de med at små skoler er bedre av følgende grunner:

- Store skoler har i motsetning til små skoler en negativ påvirkning på elever med lav sosioøkonomisk bakgrunn. Elever med høy sosioøkonomisk bakgrunn er tjent med store skoler, men effekten er veldig liten.
- Det er større og mer meningsfull deltagelse i utenomfaglige aktiviteter blant elever ved små skoler.
- Det er lavere frafallsrater, høyere oppmøterater, og flere elever som uteksamineres ved små skoler enn ved store.
- Elevers holdninger, oppførsel, og tilhørighetsfølelse er bedre ved små skoler.
- Lærerne oppgir at de er mer tilfredse, og føler et større ansvar for elevenes læring ved små skoler.

3.2.2 Nyere studier

Problemet med mange av studiene før år 2000 er at de benytter seg av enkle statistiske metoder, og de tar ikke høyde for selvseleksjonsproblemet. En rekke artikler på 2000-tallet tar opp dette problemet, og forsøker å løse det ved bruk av andre metoder, blant annet instrumentvariabel-metoden (IV).

Kuziemko (2004) bruker et paneldatasett for å kontrollere for uobserverbar heterogenitet mellom skoler. Hun generer førstedifferanse estimater av effekten av skolestørrelse på elevresultater. I tillegg utnytter hun sjokk i elevtall grunnet nye skoler, nedleggelse, og sammenslåinger i en to-steps MKM estimering, for å kontrollere for felles trender i resultater og elevtall. Resultatene viser at små skoler har bedre matematikkresultater og oppmøterater enn store skoler.

Heinesen (2004) bruker ikke skolestørrelse, men heller skoledistriktstørrelse som avhengig variabel i sin analyse. Han ser på sammenheng mellom distriktstørrelse og elevutfall etter obligatorisk skolegang i Danmark. Han benytter mikrodata på elever og deres foreldre, og estimerer logit modeller for å finne sannsynligheten for å oppnå forskjellige nivåer av utdanning. Han finner at elever som går på skoler i distrikter med mindre enn 15 000 innbyggere har mindre sjanse for å fullføre høyere utdanning senere i livet sammenlignet med elever i større distrikter.

Schneider et al. (2007) bruker to alternative metoder for å se på effektene av skolestørrelse på elevutfall. De benytter HLM og en "propensity score matching" teknikk. Ved begge modellene finner de at skolestørrelse har liten påvirkning på testresultater, men de finner andre fordeler ved små skoler. Elevene ved små skoler virker å ha større forventninger til høyere utdanning, og søker seg i større grad til universiteter (college). Begge metodene har dessverre svakheter i håndtering av seleksjon som kan gi skjeve estimater, og effektene kan derfor ikke tolkes som kausale.

Barrow et al. (2013) ser på effekten av introduksjonen av "Small School Initiative" i Chicago på elevprestasjoner. De ønsker å finne ut om elever ved små skoler har bedre resultater og større uteksamineringsrate enn lignende elever ved større skoler. De bruker IV for å kontrollere for selvseleksjonsproblemet, hvor avstand til skoler benyttes som instrumentvariabel. De finner at elever ved små skoler har mindre sannsynlighet for å slutte på skolen enn elever ved store skoler, men at de faglige prestasjoner i liten grad er bedre for disse elevene.

Schwartz et al. (2013) gjennomfører en evaluering av den storstilte skolereformen som ble gjennomført i New York City på begynnelsen av 2000-tallet, som hadde som formål å redusere størrelsen på mange av byens skoler. Identifikasjonen i analysen er den samme som i (Barrow et al., 2013). De finner at nylig opprettede små skoler har en positiv effekt

på blant annet elevers fullføringsrate, mens eldre små skoler ikke har den samme effekten. De finner også at elever ved små skoler presterer dårligere enn elever ved store skoler i engelsk, mens effekten i matematikk er mindre og ikke-signifikant. Studien er et viktig bidrag til litteraturen, fordi den belyser viktigheten av å skille mellom eldre og nyere skoler når en ser på effekten av små skoler på skolekvalitet.

3.2.3 Oppsummering

Studier som ser på sammenhengen mellom skolestørrelse og elevprestasjoner har gitt forskjellige resultater og ofte små effekter. Det kan være flere årsaker til dette. For det første benyttes forskjellige resultatmål i forskjellige studier, og det er derfor ingen grunn til at disse studiene skal komme frem til samme konklusjon. Det er ingen enighet om hva som er den beste måten å måle skolestørrelses påvirkning på elevprestasjoner. Prøveresultater blir av mange sett på som et ufullstendig mål på skolekvalitet, fordi mange relevante mål på kunnskap ikke tas med i beregningen (Heinesen, 2004). En del andre ferdigheter, som for eksempel sosiale ferdigheter, er et viktig bidrag i læringen på skolen. Dersom skoler og lærere vet at de blir evaluert på prøveresultater kan det hende de tilpasser læringen etter dette. Resultatmål som går på hvordan elevene lykkes etter fullført skolegang, som for eksempel suksess i arbeidslivet eller opptak til høyere studier, er derfor foreslått som bedre mål på skolekvalitet (Heinesen, 2004).

En annen grunn til at forskjellige studier finner forskjellig effekt av skolestørrelse kan være at de på ulik måte kontrollerer for sosioøkonomisk bakgrunn. Mange studier har funnet at økt skolestørrelse gir bedre resultater for elever av høy sosial klasse, men dårligere resultater for elever med en mer ufordelaktig bakgrunn (Lee og Smith, 1997; Page et al., 2002). Dersom ulike datasett inneholder en overvekt av en de to gruppene, vil dette derfor kunne føre til forskjellige konklusjoner.

Et annet poeng er at mange av studiene undersøker elever på forskjellige klassenivåer. Det er godt mulig at skolestørrelse påvirker elever på barneskoler ganske forskjellig fra elever på ungdomsskoler. Det er blitt gjort funn på at unge elever kan være tjent med mindre skoler, mens disse effektene ikke vil gjelde for eldre elever, som heller kan dra nytte av fordelene ved større skoler (Friedkin og Necochea, 1988). Dette kan være fordi elever må oppnå et grunnivå av akademiske ferdigheter før de kan dra nytte av de potensielle fordelene ved større skoler, og at de frem til dette tidspunktet har mer nytte av å gå på mindre skoler.

Mye av forskningen på skolestørrelse har begrenset nytte på grunn av at det typisk er brukt forholdsvis enkle metoder i litteraturen. I fravær av politiske tiltak og randomiserte eksperimenter er det vanskelig å skille ut den kausale effekten av skolestørrelse fra andre faktorer som påvirker elevutfall, som for eksempel uobserverbare egenskaper ved familien

og nærmiljøet, og skolens innsatsfaktorer. Mange av studiene på 2000-tallet tar bedre høyde for dette, og er derfor mer troverdige. Konklusjonen fra en del av disse studiene (Barrow et al., 2013; Schwartz et al., 2013) er at små skoler påvirker elevenes testresultater i liten grad, men at mindre skoler har en positiv påvirkning på elevene langs en rekke andre dimensjoner. Dette bekrefter også en hypotese om at intervensjoner siktet på eldre elever kan forbedre deres ikke-kognitive ferdigheter, men i liten grad kan påvirke deres kognitive ferdigheter.

4 Estimeringsstrategi og økonometriske utfordringer

Det å finne sammenhengen mellom skolestørrelse og elevprestasjoner er som med de fleste økonometriske sammenhenger, beheftet med en del metodologiske utfordringer. Det ideelle ville ha vært å gjennomføre et eksperiment hvor elever tilfeldig ble plassert i skoler av varierende størrelse, for så å se på forskjellene i resultater mellom elevene. Et slikt eksperiment er dessverre vanskelig å gjennomføre, både på grunn av høye kostnader, og fordi det ganske sikkert vil møte mye motstand fra forskjellige interessegrupper som foreldre, skoleaktører og politikere. Derfor må man basere studier på sammenhengen mellom skolestørrelse og resultater på skoledata hvor elevene ikke er tilfeldig fordelt til skoler, og heller forsøke å ta høyde for problemene dette medfører med statistiske verktøy.

Jeg vil i dette kapitlet gå gjennom modellen jeg vil bruke, de økonometriske utfordringene i forbindelse med analysen, samt metodene jeg har tenkt til å benytte for å ta hensyn til disse utfordringene. All teorien i dette kapitlet baserer seg på læreboka til Woolridge (2009).

4.1 Modellen

Utgangspunktet for min analyse vil være en samtidig spesifisering av en generell skoleproduktfunksjon (se delkapittel 2.2)

$$T_{is} = \alpha_0 + SS_s\beta_1 + F_{is}\beta_2 + S_s\beta_3 + \epsilon_{is}. \quad (7)$$

Her er T_{is} prøveresultat for elev i på skole s . α_0 er et konstantledd, mens SS_s er skolestørrelse på skole s . F_{is} er en vektor av elev- og familiekarakteristikk som for eksempel kjønn, alder, foreldres utdanning osv. S_s er en vektor av skolekarakteristikk som blant annet inkluderer karakteristikk ved lærere. ϵ_{is} er et stokastisk restledd.

Jeg vil begynne med å estimere sammenhengen mellom skolestørrelse og testresultater i (7) med bruk av minste kvadraters metode (MKM). MKM benyttes for å finne sammenhengen mellom forklaringsvariablene og utfallsvariabelen i en lineær modell. Sammenhengen mellom variablene finnes ved å minimere summen av kvadrerte avvik.

4.1.1 Utvalgs- og restleddsforutsetninger

Med utgangspunkt i skoleproduktfunksjonen vil jeg oppsummere en del forutsetninger som er nødvendige for at MKM skal gi forventningsrette, konsistente og effisiente estimatorene. Disse forutsetningene er som følger:

- (i) Modellen er lineær i parameterne. Denne forutsetningen er pålagt i modellen over, hvor det er inkludert både et skjæringspunkt og en helningsparameter.
- (ii) Utvalget er tilfeldig trukket ut fra populasjonen. Denne forutsetningen er oppfylt i mitt datasett, noe som diskuteres grundigere i kapittel 5.
- (iii) Det er utvalgsvariasjon i forklaringsvariabelen, som vil si at forklaringsvariabelen ikke er lik for alle observasjonene. Dette er nødvendig for at man i det hele tatt skal kunne finne noen effekter av skolestørrelse på elevprestasjoner.
- (iv) Restleddet har en forventet verdi på null, uavhengig av verdien på kontroll- og forklaringsvariablene. Dette kan skrives som

$$E(\epsilon_{is} | SS_s, F_{is}, S_s) = 0.$$

- (v) Restleddene er homoskedastiske, det vil si at variansen til restleddet er lik uavhengig av verdiene på kontroll- og forklaringsvariablene. Dette kan skrives som

$$Var(\epsilon_{is} | SS_s, F_{is}, S_s) = \sigma^2.$$

- (vi) Restleddene er uavhengige på tvers av observasjoner. Dette kan skrives som

$$E(\epsilon_{is}\epsilon_{js} | SS_s, F_{is}, S_s) = 0. \quad i \neq j$$

4.2 Brudd på eksogenitetsforutsetningen

Ved brudd på forutsetning (iv), som sier at restleddet har en forventet verdi på null uavhengig av verdien på forklaringsvariablene, vil MKM generelt ikke gi forventningsrette estimatorene. Et vanlig problem i økonometriske analyser er brudd på eksogenitetsforutsetningen. Normale årsaker til endogenitet i forklaringsvariabelen er utelatte variable, målefeil og simultanitet, hvor simultanitet i min analyse blir et spørsmål om seleksjon. Jeg vil i det videre beskrive hvert av disse problemene.

4.2.1 Utelatte variable

Dersom vi utelater en variabel fra modellen som er korrelert med både den avhengige variabelen og forklaringsvariabelen, vil det oppstå skjevhet i estimatorene. Dette kan illustreres med en situasjon hvor det er en utelatt variabel. Jeg antar at den sanne modellen er gitt ved

$$T_{is} = SS_s\beta_1 + X_{is}\beta_2 + \epsilon_{is}, \quad (8)$$

der $E(\epsilon_{is}|SS_s, X_{is}) = 0$, SS_s er skolestørrelse, og X_{is} er en variabel som påvirker elevprestasjoner. Dersom vi i stedet estimerer

$$T_{is} = SS_s\beta_1 + \epsilon_{is} \quad (9)$$

med MKM, vil det innebære at det oppstår skjevhet i det endelige utvalget gitt ved

$$E(\widehat{\beta}_1) - \beta_1 = \beta_2\widehat{\delta}, \quad (10)$$

der $\widehat{\delta}$ er koeffisienten foran SS_s i regresjonen av X_{is} mot SS_s .

I modellen med skolestørrelse og testresultater er det sannsynlig at vi kan ha utelatte variable, både på skolenivå, familienivå og på individnivå.

På skolenivå er det mulig å inkludere kontrollvariabler for veldig mye som potensielt kan ha en påvirkning på både skolestørrelse og elevresultater, som for eksempel lærerkvalitet og ressurser. Det kan ikke utelukkes at det finnes utelatte variable på skolenivå, men det er ikke først og fremst på dette nivået hvor bekymringen ligger.

På familienivå har man ofte data på foreldres utdanning, yrke og ressursituasjon, og dette kan det derfor normalt kontrolleres for. Foreldres forventninger eller hvor mye tid foreldre bruker på leksehjelp er ting det er vanskeligere å kontrollere for. Leksehjelp vil mest sannsynlig påvirke elevenes resultater, og hvis det i tillegg er en sammenheng mellom hvor mye leksehjelp elevene får og skolestørrelse, vil vi ha et utelatt variabel problem.

Også på individnivå er det vanskelig å inkludere alle nødvendige kontrollvariabler som påvirker elevprestasjoner. En stor utfordring når en skal spesifisere en produktfunksjon for kognitiv ferdigheter er at kunnskapsoppbygging er en kumulativ prosess og avhenger av medfødte evner (se delkapittel 2.2). Det er blitt vist at kunnskapsnivået ved skolestart er svært viktig (Todd og Wolpin, 2003). Det er veldig vanskelig å finne gode mål på elevers medfødte evner og dette kan potensielt skape et utelatt variabel problem. Hvis det er en typisk skolestørrelse som tiltrekker seg elever med høyt evnenivå, vil vi kunne finne en sammenheng mellom denne skolestørrelsen og gode prestasjoner som sikkert ikke har noe med størrelsen på skolen å gjøre.

4.2.2 Målefeil

Et potensielt problem i alle økonometriske analyser er at variablene kan være beheftet med målefeil. Målefeil kan defineres som et avvik mellom den variabelen vi observerer, det vil si har måling på, og den teoretisk riktige variabelen. Målefeil kan oppstå både i den avhengige variabelen og i forklaringsvariabelen. Under den vanlige forutsetningen om at målefeilen ikke er korrelert med den uobserverbare (den teoretisk riktige) forklaringsvariabelen, vil målefeil i den avhengige variabelen ikke gi opphav til skjevhet i MKM-estimatorene, men vil generelt føre til høyere varians for MKM-estimatorene. Under samme forutsetning vil målefeil i en uavhengig variabel føre til en MKM-koeffisient som er skjev mot null. Bekymringen i min analyse er rapporteringsfeil i forklaringsvariabelen, og det kan ikke utelukkes at skolestørrelse blir rapportert inn feil. Hvis dette er tilfellet, vil målefeil føre til skjeve MKM-estimatorer og gi høyere varians for estimatorene.

4.2.3 Seleksjon

En annen viktig årsak til at vi kan få endogenitet i forklaringsvariabelen er simultanitet. Dersom forklaringsvariabelen bestemmes simultant med den avhengige variabelen, vil slik endogenitet kunne oppstå. I min analyse vil dette være et spørsmål om seleksjon, både blant lærere og foreldre/elever.

Det er ikke utenkelig at flinke lærere søker seg til skoler av en viss størrelse. Dersom slik selvseleksjon av lærere ikke kontrolleres for, vil restleddet i modellen være korrelert med skolestørrelse. Hvis for eksempel flinke lærere søker seg til små skoler og dette ikke kontrolleres for, vil det underestimere en eventuell positiv effekt av skolestørrelse på testresultater. Ved å inkludere kontrollvariabler på lærerkvalitet er håpet at dette problemet forsvinner.

En større bekymring er at foreldre med vilje kan søke barna sine til skoler av en viss størrelse, ved for eksempel å bosette seg i nærheten av en slik skole. Kanskje vil små skoler virke mer appellerende på foreldre, på grunn av et tryggere miljø og en større sannsynlighet for å havne i mindre klasser hvor læreren har mer tid til hver elev. Hvis slik seleksjon i større grad er tilfellet for høyt presterende elever, vil en feilaktig kunne tilegne skolestørrelse skylden for forskjeller i prestasjoner. Den uobserverbare effekten vil inkluderes i restleddet, og estimering av modellen med MKM vil gi forventningsskjeve og inkonsistente estimatorer som vist over.

4.2.4 Kontrollvariabler og proxy-variabler

En mulig strategi for og hanskes med endogenitet å ivareta korrelasjonen mellom restledd og klassestørrelse, er å inkludere et stort antall kontrollvariabler og estimere modellen med MKM. For at denne strategien skal gi konsistente og forventningsrette estimatorer, må alle variabler som potensielt påvirker den avhengige variabelen og samtidig en av de uavhengige variablene, inkluderes i regresjonen.

Problemet med denne strategien er at det er veldig vanskelig å finne data på alle variabler som påvirker elevprestasjoner. Som nevnt i delkapittel 2.2, er det ikke lett å finne et mål på elevers medfødte evnenivå, og dette blir derfor neppe godt nok kontrollert for ved bruk av MKM. Hvor mye innsats foreldre legger i leksehjelp eller andre foreldrebidrag som påvirker barnas resultater kan det også være vanskelig å finne data på.

En vanlig brukt strategi i litteraturen er å bruke en eller flere proxy-variabler. Disse inkluderes ikke fordi de nødvendigvis har en direkte effekt på den avhengige variabelen, men fordi de er korrelert med utelatte variable og vil gjøre skjevheten på grunn av de utelatte variablene mindre. Et eksempel kan være å bruke familieinntekt som proxy-variabel for innsatsfaktorer på familienivå. Bruken av proxy-variabler må gjøres med forsiktighet fordi inkludering av disse i visse tilfeller kan føre til større skjevhet i estimatorene (Todd og Wolpin, 2003).

Selv med et stort antall kontrollvariabler vil man neppe klare å kontrollere for alle faktorer som påvirker den avhengige variabelen, og vi vet ikke med sikkerhet hvilke variabler dette er. Dette medfører at det er liten sjanse for å finne robuste og troverdige estimater av den kausale effekten av skolestørrelse på prestasjoner når man benytter denne strategien. Dette motiverer for bruken av instrumentvariabel-metoden (IV).

4.2.5 Instrumentvariabel-metoden

Instrumentvariabel-metoden tar hensyn til at restleddet er korrelert med en forklaringsvariabel i modellen. For at man skal kunne gjennomføre metoden trenger man en variabel, Z , som påvirker skolestørrelse, men som ikke påvirker prestasjonene direkte. I tillegg til strukturligningen gitt ved (7), kan vi sette opp en førstestegsrelasjon som relaterer forklaringsvariabelen i strukturligningen med instrumentvariabelen

$$SS_s = \pi_0 + Z_s\pi_1 + F_{is}\pi_2 + S_s\pi_3 + \eta_{is}. \quad (11)$$

Predikert verdi fra førstesteget benyttes som instrumentvariabel ved estimering av parametrene i strukturligningen gitt ved (7). Forutsetningene for at Z skal være en gyldig

instrumentvariabel er at

$$Cov(SS_s, Z_s) \neq 0, \quad (12)$$

og

$$Cov(\epsilon_{is}, Z_s) = 0. \quad (13)$$

(12) er relevanskriteriet, og sier at instrumentet må være korrelert med forklaringsvariabelen. (13) er eksklusjonsforutsetningen, som sier at instrumentet ikke må være korrelert med restleddet i ligningen vi ønsker å estimere, det vil si at instrumentet er eksogent.

For å finne en instrumentvariabel som påvirker skolestørrelse, men som ikke er korrelert med restleddet i ligningen av interesse, har jeg basert meg på tidligere forskning som er gjort på området. I både Barrow et al. (2013) og Schwartz et al. (2013) bruker de avstand mellom elevers hjem og skoler som instrument for skolestørrelse. Tanken er at jo større avstanden er mellom en elevs hjem og nærmeste skole som er liten, jo mindre er sannsynligheten for at eleven går på en liten skole. På samme måte ønsker jeg å bruke en instrumentvariabel som baserer seg på avstand til skoler. Siden mitt datasett inneholder mye mindre informasjon om lokasjon på elevenes bosted og lokasjon på skolene, blir jeg dessverre nødt til å gjøre det på en litt annen måte.

I spørreskjemaet som skolelederne svarer på i forbindelse med PISA-undersøkelsen, blir de spurt om størrelsen på lokalmiljøet, hvor de får fem svaralternativer. Jeg vil bruke dummyvariable for hvert av disse svaralternativene (med ett av svaralternativene som referansekategori) som instrumentvariabler i førstestegsligningen. Motivasjonen bak bruken av disse instrumentene kan relateres til avstandsinstrumentet i Barrow et al. (2013) og Schwartz et al. (2013). Når mange har kort reiseavstand til nærmeste skole, som er tilfellet i store lokalmiljøer, vil elevene i gjennomsnitt i liten grad gå på små skoler. I mindre lokalmiljøer, hvor reiseavstanden til skolene er større, er det mer sannsynlig at elevene går på små skoler.

For at lokalmiljøets størrelse skal være et godt instrument, må relevanskriteriet og eksklusjonsforutsetningen være oppfylt. Håpet er at det er nok korrelasjon mellom størrelsen på lokalmiljøet og skolestørrelse i førstesteget til å gi modellen tilstrekkelig forklaringskraft. En annen utfordring for identifikasjon er at det kan være en direkte effekt av kommune-størrelse på elevprestasjoner, og dermed brudd på eksklusjonsforutsetningen. Bruken av ugyldige instrumenter kan føre til en inkonsistent og skjev IV-estimator, som faktisk kan være forbundet med mer skjevhet enn en tilsvarende MKM-estimator (Murray, 2006).

Relevanskriteriet er testbart fordi det eksisterer observasjoner både på forklaringsvaria-

belen og instrumentvariablene. Man kan gjennomføre en F-test av nullhypotesen om at instrumentvariablene ikke er relevante mot alternativhypotesen om at de er det. En vanlig tommelfingerregel er at dersom testen gir en F-verdi over 10, er instrumentene relevante (Bound et al., 1995).

Eksklusjonsforutsetningen er ikke direkte testbar, men man kan allikevel få en indikasjon på om instrumentene er valide ved å gjennomføre en test for overidentifiserende restriksjoner. Dette krever at man har flere instrumentvariabler enn endogene forklaringsvariabler. Sargan-testen er en test på instrumentvariablenes gyldighet, og tester om de er korrelert med residualene i strukturligningen. Dette gir informasjon om instrumentene faktisk er eksogene. Problemet med testen er at den forutsetter at minst ett av instrumentene er eksogene, noe vi ikke kan vite med sikkerhet (Murray, 2006). Når jeg bruker indikatorvariabler på lokalmiljøets størrelse som instrumenter, må jeg anta at minst ett av instrumentene er eksogene for at Sargan-testen skal gi informasjon om de andres eksogenitet. Siden instrumentvariablene mine er veldig like, er det uansett naturlig å anta at dersom ett er eksogent er de andre det også, så Sargan-testen gir derfor begrenset med informasjon.

4.3 Effisiens

Ved estimering med MKM ønsker man generelt at estimatorene skal være effisiente. Estimatorene bør gi et mest mulig presist anslag på den ukjente parameteren, og denne presisjonen kan måles ved estimatorens varians. Vi har en effisient estimator hvis det ikke finnes noen andre estimatorene for samme parameter med mindre varians.

Foreløpig har kun konsekvenser av brudd på eksogenitetsforutsetningen blitt diskutert, og jeg har ikke sagt noe om brudd på de andre restleddsforutsetningene. I det videre vil jeg diskutere konsekvenser av brudd på disse forutsetningene, og hvordan jeg tar høyde for dette i min analyse.

4.3.1 Heteroskedastiske restledd

Forutsetning (v) sier at vi har homoskedastiske restledd, det vil si at vi har samme restleddsvariens for alle observasjonene. Dersom vi i stedet antar heteroskedastiske restledd, kan vi skrive

$$Var(\epsilon_{is} | SS_s, F_{is}, S_s) = \sigma_i^2, \quad (14)$$

der restleddsvariansen er forskjellig for ulike observasjoner.

Dersom vi beholder forutsetningen om at restleddet har en forventningsverdi lik null, vil MKM fortsatt gi forventningsrette og konsistente estimatorer, men heteroskedastisitet vil påvirke variansen til MKM-estimatorene. Generelt vil vi med heteroskedastiske restledd ikke kunne bruke de normale formlene for varians og standardavvik, og de fleste tester, som for eksempel t-test og F-test, vil ikke være gyldige hvis vi ikke tar hensyn til heteroskedastisitet.

4.3.2 Korrelerte restledd

Forutsetning (vi) sier at restleddene må være uavhengige på tvers av observasjoner. Dersom vi i stedet antar

$$E(\epsilon_{is}\epsilon_{js}|SS_s, F_{is}, S_s) \neq 0, \quad i \neq k \quad (15)$$

vil dette igjen få konsekvenser for variansen til MKM-estimatorene. Også med korrelerte restledd vil de vanlige formlene for varians og standardavvik være uegnet, og i tillegg vil MKM-estimatorene ikke være effisiente.

4.3.3 Cluster robuste standardfeil

I min analyse er skolestørrelse lik for alle elever på samme skole. Høyst sannsynlig vil derfor restleddene være korrelert innenfor skoler, og det er derfor nødvendig å ta høyde for dette. Det kan heller ikke antas at restleddene er homoskedastiske. Siden jeg bruker statistikkprogrammet Stata for å kjøre regresjonene i min analyse, velger jeg å løse dette problemet ved å benytte meg av cluster robuste standardfeil. Dette vil korrigere for feilspesifikasjon i restleddene i modellen. Ved å bruke skoler som cluster, løser jeg på homoskedastisitetsforutsetningen, samtidig som at jeg tar høyde for variasjon mellom skolene.

5 Beskrivelse av data

I dette kapitlet vil jeg gi en grundig beskrivelse av datasettet som danner grunnlaget for min analyse. Datasettet består av data fra PISA-undersøkelsene i 2006, 2009 og 2012, og inkluderer Danmark, Finland, Norge og Sverige.

Jeg starter med en beskrivelse av gjennomføringen, formålet, og historien til PISA-undersøkelsen. Deretter beskriver jeg den avhengige variabelen, forklaringsvariabelen, og kontrollvariablene. Til slutt forklarer jeg hvordan jeg forholder meg til manglende data.

5.1 PISA-undersøkelsen

PISA (Programme for International Student Assessment) er en internasjonal undersøkelse som gjennomføres av OECD hvert tredje år i en rekke land. Hensikten med undersøkelsen er å kartlegge kompetansen til elever som er i avslutningen av den obligatoriske skolegangen.

Formålet med undersøkelsen er ikke først og fremst å måle i hvilken grad elevene har kompetanse på pensum i de respektive landene, men heller i hvilken grad elevenes kunnskap og ferdigheter gjør dem rustet til videre studier og arbeidsliv. I den norske hovedrapporten i forbindelse med PISA 2012 står det: “PISA-undersøkelsen tar ikke utgangspunkt i landenes læreplaner, men tar i hovedsak sikte på å måle evne til aktivt å bruke kunnskaper og erfaringer i aktuelle situasjoner” (Kjærnsli og Olsen, 2013). Elevene testes i matematikk, naturfag og lesing, men ett av fagområdene testes grundigere hver gang. I 2012 var matematikk det vektlagte fagområdet.

I tillegg til de faglige oppgavesettene elevene må gjennomføre, svarer de på et spørreskjema som gir informasjon om elevenes kjønn, alder, sosioøkonomiske bakgrunn, motivasjon og en rekke andre forhold. Også skoleledere svarer på en spørreundersøkelse i forbindelse med PISA, og denne gir informasjon om strukturelle forhold ved skolene, som blant annet størrelse, kjønnsfordeling, antall lærere med utdanning og størrelsen på lokalmiljøet hvor skolen er lokalisert. Spørreskjemaene avdekker en del forhold ved elevene og skolene som er av interesse for forskere, lærere og politikere, og som gjør det mulig å gjennomføre endringer der hvor problemer avdekkes.

Antall deltagende land varierer litt fra gang til gang, men i 2012 deltok 65 land og 510 000 elever. Av disse var 25 700 fra Danmark, Finland, Norge eller Sverige, og representerte i overkant av 1000 skoler. Hovedvekten av de deltagende landene er OECD land eller relativt rike land.

Elevene som deltar i undersøkelsen er 15 år gamle, og testene gjennomføres uavhengig av

klassetrinn. Det vil derfor være veldig små forskjeller i alder mellom elevene, men det vil være noe forskjell i klassetrinn dersom noen elever har begynt for tidlig eller for sent på skolen. Uttrekningen av skoler til undersøkelsen foregår på en slik måte at sannsynligheten for å bli trukket ut er proporsjonal med skolestørrelse. På denne måten er det omtrent like stor sannsynlighet for at hver elev i et land trekkes ut, og man ender opp med et representativt utvalg.

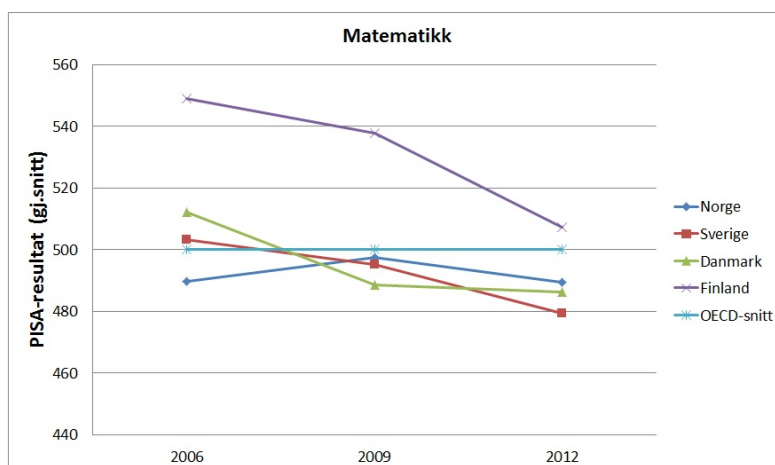
Et potensielt problem i undersøkelsene er fritak av elever. Dersom elever med spesielle behov går på spesialskoler i et land, mens de er integrert i det offentlige skolesystemet i et annet, vil det være vanskelig å sammenligne resultater på tvers av land. Det finnes derfor et strengt regelverk på hvilke elever som kan fritas fra undersøkelsene. De som kan fritas er elever med en fysisk funksjonshemming som gjør at de ikke kan gjennomføre prøven, elever med kognitive, psykiske og/eller emosjonelle vansker, eller elever med begrensede språkkunnskaper i testspråket. Antall fritatte elever har økt i flere av de deltagende landene fra 2000 til 2012, inkludert Norge og deler av Norden, og dette er et problem fordi fritaksprosenten er ujevn på tvers av land.

5.2 Elevutfall

De avhengige variablene jeg benytter i analysen er gjennomsnittresultatet i henholdsvis matematikk, lesing og naturfag. For hvert av fagene er det i datasettet oppgitt fem sannsynlige verdier (plausible values) for hver elev, hvor hvert av de fem nivåene representerer en gruppe av PISA-spørsmål. Nivå 1 er det letteste nivået, og spørsmålene krever kun grunnleggende ferdigheter for å kunne løses. For hvert nivå øker vanskelighetsgraden. Hver elev gjennomfører kun en liten del av det totale oppgavesettet, fordi det er begrensninger på hvor mye tid som er til rådighet, og dette medfører en grad av usikkerhet rundt prestasjonsmålene. Bruken av sannsynlige verdier er et forsøk på å løse dette problemet, og er en metode hvor man bruker flere verdier som representerer den sannsynlige fordelingen av elevens ferdigheter. Jeg benytter gjennomsnittet av de fem sannsynlige verdiene i hvert fag som avhengig variabel. Resultatene på testene er skalert til å ha et gjennomsnitt på 500 poeng med et standardavvik på 100.

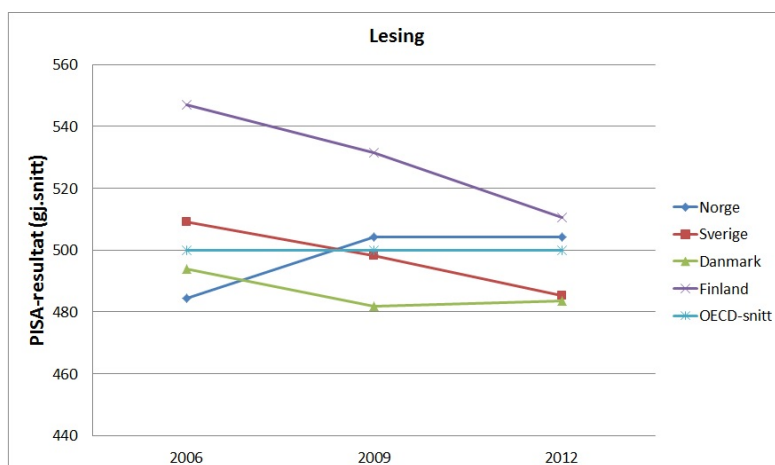
Figur 2, 3 og 4 viser utviklingen i de tre testfagene fra 2006 til 2012 for hvert av de nordiske landene. I Figur 2 ser vi utviklingen i matematikkresultater. Norge har ligget noe under OECD-gjennomsnittet på alle de tre undersøkelsene, og hadde en liten økning fra 2006 til 2009, men en reduksjon fra 2009 til 2012. Både Sverige og Danmark lå over OECD-gjennomsnittet i 2006, men har sett en reduksjon i resultater siden den gang. Finland har ligget langt over både OECD-gjennomsnittet og sine nordiske naboer i alle de tre årene, men er det landet som har hatt den mest drastiske nedgangen i resultater siden 2006.

Figur 2: Utvikling i matematikkresultater i Norden



Figur 3 viser utviklingen i leseresultater. Her ser vi at Norge, som lå under OECD-gjennomsnittet i 2006, så en økning til over snittet i 2009 og har opprettholdt denne posisjonen til 2012. Sverige har hatt en jevn nedgang fra over OECD-gjennomsnittet i 2006 til å ligge et stykke under snittet i 2012. Danmark har ligget under snittet hele veien og har i likhet med Sverige sett en nedgang i resultater siden 2006. Utviklingen for Finland er ganske lik i lesing som i matematikk, fra langt over OECD-gjennomsnittet i 2006, til kun litt over i 2012.

Figur 3: Utvikling i leseresultater i Norden

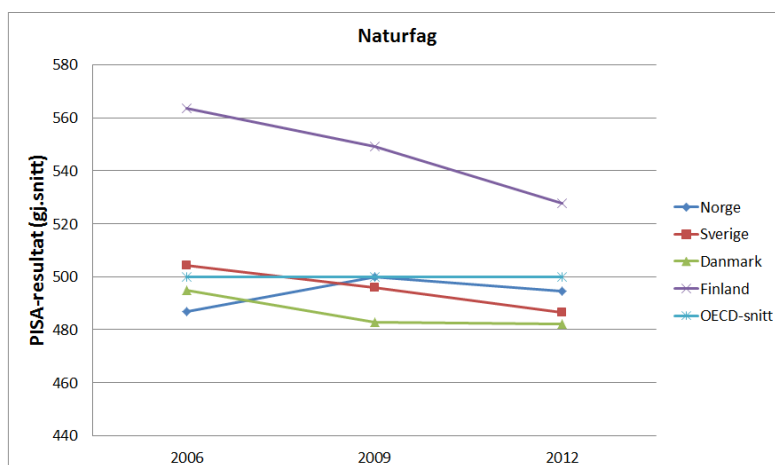


Figur 4 viser utviklingen i naturfagsresultater. Vi ser her mye av den samme utviklingen i resultater for de nordiske landene som for i lesing.

5.3 Skolestørrelse

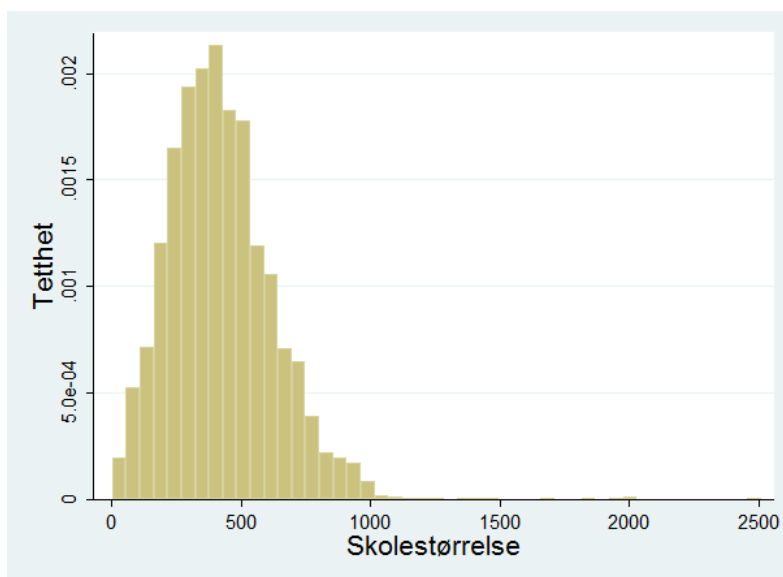
Forklaringsvariabelen jeg benytter i min analyse er skolestørrelse. Datasettet inneholder tre skoler som har en størrelse på over 3000 elever. Jeg mistenker at dette kan skyldes en

Figur 4: Utvikling i naturfagsresultater i Norden



rapporteringsfeil, og velger derfor å fjerne disse observasjonene. Siden dette kun gjelder tre skoler og totalt 12 elever, skaper ikke dette noen problemer for analysen. I utvalget har skolestørrelse en gjennomsnittsverdi på 424,61 med en standardfeil på 203,38. Minste skole har 3 elever, mens største skole har 2512 elever. Figur 5 viser histogram over skolestørrelse for hele utvalget. De fleste skolene har en størrelse på mellom 0 og 1000 elever, og hovedvekten ligger rundt 500. Svært få skoler har over 1000 elever.

Figur 5: Histogram over skolestørrelse

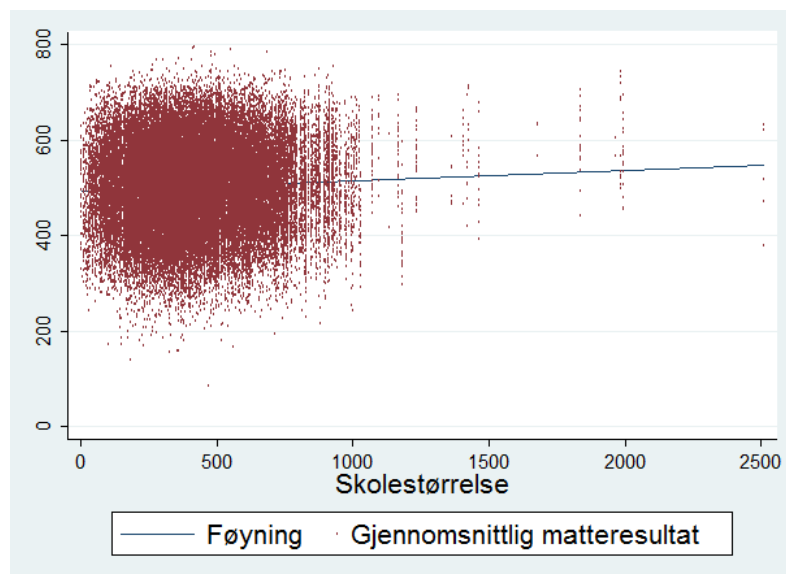


Figur 6 viser spredningsplottet mellom skolestørrelse og gjennomsnittsresultater i matematikk, inkludert føyningen, det vil si den linjen som passer best til observasjonene. Hvert av punktene i spredningsplottet representerer en dataobservasjon, og denne fremstillingen gir en indikasjon på hva slags sammenheng det er mellom skolestørrelse og testresultater i matematikk. Spredningsplottet viser en svak positiv sammenheng mellom skolestørrelse og matematikkresultater. De fleste skolene har en skolestørrelse på mellom 0 og 1000

elever, og i dette intervallet ser vi at spredningen i resultater er stor. Blant skolene med størrelse over 1000 elever ser vi at resultatene jevnt over synes å være høye, det er få observasjoner som ligger langt under gjennomsnittet, og det er mange observasjoner som ligger på gjennomsnittet eller over. Figur 7 i appendiks viser samme spredningsplott, men uten skoler med over 1000 elever. Vi ser her en mer jevn fordeling av punktene rundt føyningen, og den positive sammenhengen mellom skolestørrelse og testresultater virker å være litt svakere. Fra figur 6 kan det virke som om elever ved store skoler presterer bedre enn elever ved små skoler, men vi kan ikke konkludere med at det er skolestørrelsen i seg selv som er årsaken til sammenhengen ut i fra denne enkle fremstillingen.

I tabell A2 i appendiks presenteres gjennomsnittlig skolestørrelse for de fire nordiske landene på de forskjellige PISA-undersøkelsene. Vi ser her at gjennomsnittlig skolestørrelse er ganske lik for Danmark, Finland og Sverige over de tre årene, og ligger jevnt mellom 400 og 500 elever med noe variasjon fra år til år. I Norge er gjennomsnittlig skolestørrelse en del mindre enn i de andre landene, og ligger under 350 elever i alle de tre årene. En del av forklaringen på dette er nok at de største skolene i Norge virker å være ganske mye mindre enn i de andre landene.

Figur 6: Spredningsplott med skolestørrelse og matematikkresultater



5.4 Kontrollvariabler

I dette delkapittelet vil jeg beskrive kontrollvariablene som inkluderes i analysen. Siden de originale PISA-datasettene inneholder store mengder variabler, må det gjøres et valg på hva som skal inkluderes i modellen. Jeg har valgt å bruke de samme kontrollvariablene som Hanushek et al. (2013) bruker i sin undersøkelse. De benytter et paneldatasett bestående av de tre PISA-undersøkelsene fra 2000 til 2009 for å se på sammenhengen mellom skolers

selvstendighet i egne beslutninger og elevprestasjoner, og denne undersøkelsen er derfor på mange måter sammenlignbar med min.

Hanushek et al. (2013) bruker kontroller på tre nivåer; individ/familie-, skole-, og landnivå. Jeg benytter kun kontrollvariablene på individ/familie- og skolenivå, og inkluderer heller indikatorvariabler for land i regresjonene.

I det videre vil jeg starte med en beskrivelse av kontrollvariabler på individnivå, før jeg beskriver kontrollvariablene på skolenivå. Deskriptiv statistikk for utvalget som benyttes i analysen er rapportert i tabell A.1 i appendiks. Det henvises til denne tabellen for tall som nevnes i dette delkapittelet.

5.4.1 Elev- og familiekarakteristikk

Jente er en dummyvariabel lik én dersom eleven er jente. I utvalget er det en jenteandel på 0,499, det vil si at det er omtrent like mange jenter som gutter i utvalget.

Alder er en variabel som beskriver elevenes alder ved gjennomføringen av testen. Gjennomsnittsalderen i utvalget er 15,74 med en standardfeil på 0,29. Yngste elev er 15,17 år og eldste elev er 16,33 år.

Immigrasjonsbakgrunn kategoriseres i tre dummyvariabler som indikerer om elevene er etnisk nordiske, første- eller andregenerasjonsinnvandrere. I utvalget er 7,2 prosent av elevene førstegenerasjonsinnvandrere, og 4,7 prosent er andregenerasjonsinnvandrere.

Språk i hjemmet beskrives ved tre dummyvariabler som indikerer om testspråket er et annet enn det språket som snakkes i elevenes hjem. 8,1 prosent av elevene snakker en annen nasjonal dialekt, 1,3 prosent snakker et annet språk, mens de resterende elevene snakker samme språk som testspråket i hjemmet.

Foreldres utdanning kategoriseres i seks dummyvariabler som beskriver foreldrenes utdanningsnivå, og går fra ingen utdanning til universitetsutdanning. 61,3 av elevene har en mor med en eller annen form for universitetsutdanning, mens 52,2 prosent av elevene har en far med universitetsutdanning i utvalget. Bare 1,4 prosent av elevene har en mor uten noe utdanning, og tilsvarende har 1,4 prosent av elevene en far uten noe utdanning.

Bøker i hjemmet består av seks dummyvariable som sier noe om antallet bøker elevene har i hjemmet, og inkluderes som en proxy-variabel på sosioøkonomisk bakgrunn. Den største andelen av elevene har mellom 26 og 100 bøker i hjemmet, mens under 10 prosent av elevene sier at de har mellom 0 og 10 bøker i hjemmet eller over 500 bøker i hjemmet.

Hanushek et al. (2013) inkluderer indikatorer for foreldres yrkestilhørighet som kontrollvariabler i sin analyse. I mitt datasett finnes denne informasjonen kun for årene 2006 og

2009, og mangler derfor for over 40 % av observasjonene. Av denne grunn ser jeg meg nødt til å utelate disse variablene i min analyse.

5.4.2 Skolekarakteristikk

Privatskole er en dummyvariabel lik én dersom skolen er privat drevet. I utvalget er 9,2 prosent av skolene private. De fleste skolene er altså offentlige, noe som gir god mening med tanke på at utvalget består av nordiske skoler, hvor offentlige skoler er mest utbredt.

Jeg inkluderer også en variabel som beskriver hvor stor andel av skolenes inntekter som er finansiert av det offentlige. Gjennomsnittet i utvalget er på 98 prosent.

Andel fullt sertifiserte lærere forteller hvor stor andel av lærerne på skolene som er sertifiserte, som i nordisk sammenheng betyr at de har lærerutdanning. Gjennomsnittet i utvalget er på 89,6 prosent.

Mangel på lærere i de tre fagene kategoriseres i fire indikatorvariabler som sier noe om i hvilken grad skolen føler de mangler lærere. For alle fagene svarer de fleste skolelederne at de ikke mangler lærere i det hele tatt (mellom 65 og 80 prosent), eller at mangelen er veldig liten.

Skolens beliggenhet kategoriseres i fem dummyvariabler; lite sted med under 3 000 innbyggere, tettsted eller by med mellom 3 000 og 15 000 innbyggere, tettsted eller by med mellom 15 000 og 100 000 innbyggere, by med mellom 100 000 og 1 000 000 innbyggere, eller storby med over 1 000 000 innbyggere. Over 40 prosent av skolene ligger på et sted med under 15 000 innbyggere, mens flertallet av skolene er lokalisert på et sted med mellom 15 000 og 100 000 elever. I mitt utvalg er det kun København og Stockholm som er storbyer med mer enn 1 000 000 innbyggere, og dette forklarer hvorfor så få skoler har beliggenhet i storbyer.

5.4.3 Andre kontrollvariabler

Ved siden av variabler på elev- og skolenivå, inkluderer jeg dummyvariabler for land og år som kontrollvariabler. Jeg setter Norge som referansekategori, og inkluderer dummyvariabler for Danmark, Finland og Sverige. I tillegg setter jeg 2006 som referansekategori for år, og inkluderer dummyvariabler for 2009 og 2012. Inkluderingen av tidsdummyer og dummyer for land gjøres for å kontrollere for felles påvirkninger hvert år og i hvert land.

5.5 Manglende data

Som med de fleste datasett av en viss størrelsesorden, er også mitt datasett beheftet med manglende data på en del observasjoner. Når det mangler data for en observasjon på enten den avhengige eller en av de uavhengige variablene, kan ikke denne observasjonen benyttes i standard regresjonsanalyse. Så lenge manglende data er riktig registrert i datasettet, vil Stata holde styr på dette og ignorere observasjoner med manglende informasjon.

Hvordan man skal forholde seg til manglende data avhenger av årsaken til at dataen mangler. Dersom det er helt tilfeldig hvilke data som mangler vil dette kun føre til at størrelsen på det tilfeldige utvalget reduseres, noe som gir mindre presise estimatorer, men ingen skjevhet (Woolridge, 2009). Manglende data er mer problematisk hvis det resulterer i ikke-tilfeldig utvalg fra populasjonen. I mitt datasett vil dette være et problem dersom manglende data er mer vanlig for noen type elever eller skoler enn for andre. Hvis det for eksempel er mer manglende data på de lavt presterende elevene i utvalget enn på de høyt presterende, vil ikke utvalget være representativt for populasjonen.

I tabell A.1 i appendiks er den deskriptive statistikken for variablene i min analyse oppsummert, og i siste kolonne vises det hvor stor andel av observasjonene som har manglende data på den aktuelle variabelen. En del av kontrollvariablene mangler data på ganske mange av observasjonene, og det er derfor behov for å ta hensyn til dette i analysen. Jeg løser dette ved å inkludere indikatorvariabler som er lik én hvis informasjonen mangler for hver av variablene som har manglende data. På denne måten mister jeg ingen av observasjonene når jeg kjører regresjoner, men det resulterer i noe høyere standardavvik.

Det er også en del manglende data på skolestørrelse (litt over 5% av observasjonene). Siden skolestørrelse er forklaringsvariabelen av interesse er det spesielt viktig å unngå målefeil i denne. Jeg velger derfor å droppe observasjoner med manglende data på skolestørrelse i analysen. Dette vil føre til at jeg får færre observasjoner i regresjonene og dermed høyere standardavvik, men ellers vil det ikke medføre noen problemer.

6 Resultater

I dette kapitlet vil jeg legge frem resultatene fra min analyse. Utgangspunktet for analysen vil være minste kvadraters metode (MKM) på modellen som beskrevet i kapittel 4. Jeg vil starte med MKM av en helt enkel modell som relaterer testresultater med skolestørrelse, før jeg øker antall kontrollvariabler for å se om resultatene er robuste for inkluderingen av disse.

Som tidligere nevnt vil jeg neppe klare å kontrollere for alle faktorer som påvirker elevresultatene med MKM, og resultatene fra denne analysen må derfor tolkes med forsiktighet. Neste steg blir derfor å presentere resultatene fra instrumentvariabel-analysen (IV).

Jeg vil så kjøre de samme regresjonene på delutvalg av datasettet for å se om det eksisterer heterogene effekter. Tidligere forskning har vist at skolestørrelse har større effekt på noen elevgrupper enn andre, og det er derfor av interesse å se om de samme effektene er til stede i mitt utvalg.

Siden jeg kjører ganske mange regresjoner, og jeg må forholde meg til testresultater i tre forskjellige fag, vil jeg ikke presentere alle resultatene i dette kapitlet. Jeg vil som en hovedregel presentere og kommentere regresjonene på matematikkresultater grundig, mens regresjoner på lese- og naturfagsresultater kommenteres mindre grundig og med resultatene inkludert i appendiks. Siden skolestørrelse er interessevariabelen vil jeg stort sett kun inkludere koeffisienter for denne i tabellene, men jeg vil rapportere om hva slags kontrollvariabler som inkluderes.

6.1 Minste kvadraters metode

6.1.1 Matematikk

I tabell 1 presenteres resultatene fra MKM-regresjoner av skolestørrelse på matematikkresultater med forskjellig antall kontrollvariabler. I modellen i kolonne (1) er ingen kontrollvariabler inkludert, og koeffisienten for skolestørrelse er her lik 0,022. Koeffisienten er signifikant på ettprosentnivå, men kan ikke antas å være forventningsrett og konsistent.

I kolonne (2) til (4) inkluderes gradvis flere kontrollvariabler. Dette synes å ha liten påvirkning på koeffisienten for skolestørrelse. Modellen i kolonne (4) hvor alle kontrollvariabler er inkludert, har samme koeffisient på skolestørrelse som modellen uten kontrollvariabler, og begge koeffisientene er signifikante på ettprosentnivå. Modellen i kolonne (4) har nødvendigvis en betydelig høyere forklaringskraft, men dette er også den eneste forskjellen mellom de to modellene.

For å sette koeffisientstørrelsen på 0,022 i perspektiv kan man se for seg at vi øker størrelsen på en vilkårlig skole med ett standardavvik, som utgjør omtrent 203 elever. Dette vil gi en økning i gjennomsnittresultatet i matematikk på 4,5 poeng, som utgjør omtrent 5,2 prosent av et standardavvik. Effekten er med andre ord ikke veldig stor, men det er interessant at det er en signifikant effekt til stede.

Tabell 1: MKM-estimering med matematikkresultater som avhengig variabel

	(1)	(2)	(3)	(4)
Skolestørrelse	0,022** (0,004)	0,028** (0,003)	0,025** (0,003)	0,022** (0,003)
Kontroller for elevkarakteristikk	Nei	Ja	Ja	Ja
Kontroller for skolekarakteristikk	Nei	Nei	Ja	Ja
Dummyer for land og år	Nei	Nei	Nei	Ja
R^2	0,003	0,245	0,250	0,278
N(Elever)	61543	61543	61543	61543
Cluster(Skoler)	2528	2528	2528	2528

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i matematikk. Estimert med minste kvadraters metode. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler. Full modell finnes i tabell A5 i appendiks

6.1.2 Lesing og naturfag

I tabell A3 og A4 i appendiks presenteres resultatene fra MKM-regresjoner av skolestørrelse på testresultater i henholdsvis lesing og naturfag. Effekten av skolestørrelse er for begge fagene veldig lik effekten av skolestørrelse på matematikkresultater. Når alle kontrollvariabler inkluderes er koeffisienten for skolestørrelse på leseresultater lik 0,027, mens den er 0,022 for naturfag. Begge effektene er signifikante på ettprosentnivå.

6.1.3 Effekt av kontrollvariabler

Jeg vil i det videre gå gjennom effektene av de forskjellige kontrollvariablene på testresultater. Det er interessant å se hvilken påvirkning kontrollvariabler på elev- og skolenivå har på elevprestasjoner. Resultatene fra MKM-regresjoner av skolestørrelse på testresultater i de tre fagene er presentert i tabell A5 i appendiks, hvor koeffisientene for de forskjellige kontrollvariablene er inkludert.

Elevenes kjønn har en signifikant effekt på resultatene i alle de tre fagene. I matematikk og naturfag presterer gutter bedre enn jenter, og forskjellen er på henholdsvis 11,6 og 4,3

poeng. I lesing presterer jenter derimot bedre enn gutter, og forskjellen er på så mye som 38,4 poeng. Dette samsvarer med den normale oppfatningen om at gutter presterer bedre enn jenter i realfag, mens jenter presterer bedre i lesing.

Elevenes alder viser seg å ha en effekt på resultatene. Som tidligere nevnt er alle elevene som testes født i samme år. Det er allikevel en forskjell i når på året elevene er født, og også hvilket klassetrinn elevene går på (selv om de fleste går på samme klassetrinn). En økning på ett år i alder gir en signifikant forskjell i resultater på mellom 10,5 og 11,5 poeng i alle de tre fagene.

Også immigrasjonsbakgrunn virker å påvirke testresultatene. Innvandrere presterer signifikant dårligere enn etnisk nordiske elever i alle testfagene. Forskjellen i poeng er stor, og ligger mellom 13 og 41 poeng avhengig av fag og innvandringsbakgrunn. Videre ser vi at andregenerasjonsinnvandrere presterer bedre enn førstegenerasjonsinnvandrere.

Elever som snakker en annen dialekt eller et annet språk enn testspråket i hjemmet, presterer signifikant dårligere på testene enn elever som snakker samme språk som testspråket. Videre presterer elever som snakker en annen nasjonal dialekt dårligere enn de som snakker et annet språk, og denne forskjellen er til stede i alle de tre testfagene.

Som forventet har foreldres utdanning en stor påvirkning på elevenes testresultater. I alle de tre fagene ser vi at elevenes testresultater øker med foreldrenes utdanningsnivå, og alle koeffisientene er signifikante på ettprosentnivå. For eksempel vil en elev som har en far med minst en bachelorgrad, prestere 36,6 poeng bedre på PISA-testen i matematikk enn en elev med en far uten utdanning, eller 17,6 poeng bedre enn en elev som har en far med kun grunnskoleutdanning. Både mors og fars utdanning påvirker testresultatene signifikant, men fars utdanning har den største effekten.

Antall bøker i hjemmet brukes som en proxy på sosioøkonomisk bakgrunn, og viser seg å ha en stor effekt på hvordan elevene presterer på testene. I alle fagene ser vi en økning i testresultater med antall bøker, og koeffisientene er store og signifikante. En elev som har over 500 bøker i hjemmet vil prestere 84 poeng bedre i matematikk enn en elev som har mellom 0 og 10 bøker i hjemmet.

Elever ved private skoler presterer bedre enn elever ved offentlige skoler i alle fag. Effekten er størst i lesing, hvor elever ved private skoler presterer 14,5 poeng bedre enn elever ved offentlige skoler. Effekten er signifikant på ettprosentnivå.

Hvor stor andel av skolens finansiering som kommer fra det offentlige og andelen lærere med lærerutdanning har ingen signifikant effekt på testresultater i noen av fagene.

Effektene av lærermangel i de tre fagene varierer veldig, og avhenger av i hvilken grad skoleledelsen oppgir at det er mangel. I matematikk og naturfag er effektene av lærermangel relativt små, og ikke signifikante i alle tilfeller. I lesing presterer elever ved skoler

som oppgir at de har stor lærermangel 19,3 poeng bedre enn elever ved skoler som oppgir at de ikke har lærermangel i det hele tatt. Dette må sies å være en overraskende effekt.

Indikatorvariablene for land og år samsvarer i stor grad med den deskriptive statistikken i kapittel 5. Finland presterer mye bedre enn de andre landene i alle fag, mens forskjellene mellom de tre andre landene er mindre. Testresultatene har svekket seg siden 2006, og særlig mye i matematikk.

Vi ser at det er variablene på elev- og familiekarakteristikk som har størst påvirkning på elevenes testresultater, noe som ikke er en stor overraskelse. Elevers kjønn, alder, immigrasjonsbakgrunn, samt foreldres utdanning og sosioøkonomisk bakgrunn er alle forventet å ha en relativt stor påvirkning på resultater, noe som bekreftes i MKM-regresjonene.

6.1.4 Ikke-lineær funksjon

Jeg har i utgangspunktet antatt at det er en lineær sammenheng mellom testresultater og skolestørrelse, men dette er ikke nødvendigvis tilfellet. Det kan derfor være interessant å inkludere et kvadratisk ledd for skolestørrelse i modellen for å se om det finnes en ikke-lineær sammenheng mellom de to variablene.

I tabell 2 presenteres resultatene fra MKM-regresjoner av skolestørrelse på matematikkresultater, hvor jeg har inkludert et kvadratisk ledd for skolestørrelse. Bortsett fra det kvadratiske leddet, er modellene i kolonne (1) til (4) identiske med de i tabell 1. Det kvadratiske leddet er signifikant i modellen uten kontrollvariabler, men blir ikke-signifikant når kontrollvariabler inkluderes. Ved første øyekast kan den kvadratiske effekten virke liten, men ved en stor økning i skolestørrelse vil effekten av det kvadratiske leddet bli stor. I figur 8 i appendiks presenteres sammenhengen mellom matematikkresultater og skolestørrelse fra kolonne (4) i tabell 2. Her vises skolestørrelse på x-aksen og resultat i poeng på y-aksen. Grafen er konveks, og betydningen av det kvadratiske leddet på resultatene er økende i skolestørrelse. Siden ikke alle koeffisientene er signifikante, kan det ikke legges for mye i disse resultatene, men det gir allikevel en indikasjon på at sammenhengen mellom skolestørrelse og testresultater kan være ikke-lineær og økende i skolestørrelse.

6.1.5 Skjeve estimater?

MKM-resultatene jeg har presentert i dette delkapittelet tyder på at det er en positiv signifikant effekt av skolestørrelse på elevers resultater på PISA-undersøkelsene. Spørsmålet blir da om disse effektene kan tolkes som kausale, eller om det er grunner til å mistenke skjevheter i koeffisientene.

Tabell 2: MKM-estimering med matematikkresultater som avhengig variabel, inkludert ledd for kvadrert skolestørrelse

	(1)	(2)	(3)	(4)
Skolestørrelse	0,006 (0,009)	0,022** (0,007)	0,019* (0,008)	0,015† (0,008)
Kvadrert (skolestørrelse/10)	0,0014* (0,0007)	0,0006 (0,0006)	0,0005 (0,0006)	0,0008 (0,0006)
Kontroller for elevkarakteristikk	Nei	Ja	Ja	Ja
Kontroller for skolekarakteristikk	Nei	Nei	Ja	Ja
Dummys for land og år	Nei	Nei	Nei	Ja
R^2	0,003	0,245	0,250	0,253
N(Elever)	61543	61543	61543	61543
Cluster(Skoler)	2528	2528	2528	2528

Signifikansnivå : † : 10% * : 5% ** : 1%

Bortsett fra som indikert er modellspesifikasjonene de samme som i tabell 1. Avhengig variabel er testresultater i matematikk. Estimert med minste kvadraters metode. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i parenteser er korrigert for å ta høyde for clustering innen skoler

I delkapittel 4.2 beskrev jeg hvordan brudd på eksogenitetsforutsetningen kan gi forventningskjevne MKM-estimatorer på grunn av utelatte variable, målefeil og seleksjon. Jeg sa ingenting om forventet retning på skjevheten i modellen. Dette er generelt ikke enkelt å si noe om, siden forskjellige effekter kan trekke i forskjellig retning, men det er allikevel mulig å gjøre seg opp en mening om forventet retning på skjevheten basert på intuisjon.

Den største bekymringen i min analyse er selvseleksjon blant foreldre. Jeg mistenker at foreldre som bryr seg mye om barnas prestasjoner vil forsøke å velge mindre skoler for sine barn hvis de har muligheten til det, for eksempel ved å bosette seg i nærheten av slike skoler. Små skoler virker ganske sikkert tryggere på foreldre, og sannsynligheten for at elevene havner i små klasser er større, og dette er faktorer jeg tror foreldre legger vekt på i valg av skole. Dersom det er egenskaper ved de elevene som systematisk plasseres i skoler av en viss størrelse som ikke kontrolleres for, vil vi få skjeve koeffisienter. Hvis disse elevene for eksempel har høyere medfødt evnenivå enn gjennomsnittet, eller har foreldre som har høyere forventninger og bidrar mer til leksehjelp, og vi antar at de velger seg til små skoler, vil vi få en skjevhet mot null i MKM-estimatoren for skolestørrelse. Effekten av økt skolestørrelse drives da til en viss grad av at flinke elever bevisst har blitt plassert i mindre skoler.

Seleksjon blant lærere er et annet potensielt problem. Jeg ser for meg at også flinke lærere søker seg til små skoler. Mulighetene for mer ansvar er nok større på små skoler, og sannsynligheten for å undervise i mindre klasser er også større. Samtidig er små skoler

forbundet med mindre atferdsproblemer og vold enn store skoler. På den annen side kan det tenkes at flinke lærere søker seg til store skoler fordi det er større muligheter for å få undervise i fag hvor de har spesialisering. Ved å inkludere kontrollvariabler på lærerkvalitet i analysen, er målet at jeg klarer å ta høyde for seleksjonsproblemet. Hvis det ikke kontrolleres godt nok for lærerkvalitet, og flinke lærere søker seg til små skoler, vil i så fall også dette bidra til at MKM-estimatoren for skolestørrelse blir skjev mot null.

Det kan også være et problem at økt skolestørrelse fører til økt frafall og repetering av klasser, noe som Barrow et al. (2013) finner i sin undersøkelse. Elever som har falt fra, og som mest sannsynlig er lavt presterende elever, vil ikke testes på prøvene. Dette vil føre til at snittet blant elevene som er igjen på de store skolene vil trekkes opp i forhold til om ingen elever hadde sluttet, og MKM-estimatoren for skolestørrelse blir skjev oppover. Siden PISA tester ungdomsskoleelever, skal ikke dette være et problem i min undersøkelse.

Fordi det er mistanke om skjevhet, kan ikke estimatorene jeg har funnet med MKM tolkes som kausale. Det blir av den grunn nødvendig å benytte alternative metoder. I neste delkapittel presenteres resultatene fra regresjoner med instrumentvariabel-metoden.

6.2 Instrumentvariabel-estimering

6.2.1 Matematikk

I tabell 3 presenteres resultatene fra IV-regresjonen av skolestørrelse på matematikkresultater. I tabellen har jeg inkludert koeffisientene fra redusert form regresjonen av instrumentene på matematikkresultater, koeffisientene fra førstesteget, og IV-koeffisienten av skolestørrelse på matematikkresultater i andresteget. Resultatene fra redusert form regresjonen i første kolonne viser at det er en signifikant effekt av lokalmiljøets størrelse på matematikkresultater. Det virker altså som at det kan være en sammenheng mellom instrumentvariablene og den avhengige variabelen, noe som er problematisk.

Instrumentene i førstesteget viser seg å være svært signifikante. Det er også betryggende at effekten av lokalmiljøets betydning på skolestørrelse er økende med størrelsen på lokalmiljøet, som er selve argumentet bak bruken av disse instrumentene. En F-verdi på nesten 162 i førstesteget tyder på at instrumentene er relevante, og denne verdien er langt over tommelfingerregelen på 10 som ofte brukes i litteraturen. P-verdien i Sargan-testen medfører at vi ikke kan forkaste nullhypotesen om at instrumentene er gyldige. Selv om det er betryggende at nullhypotesen ikke forkastes, betyr ikke dette at vi kan konkludere med at instrumentene er valide.

Koeffisienten på skolestørrelse i andresteget er på 0,045, og signifikant på ettprosentnivå. Effekten er over dobbelt så stor som ved MKM med fullt sett av kontrollvariabler. Koef-

fisientstørrelsen på 0,045 betyr at en økning i skolestørrelse med ett standardavvik vil gi en økning i gjennomsnittresultatet i matematikk på 9,1 poeng, noe som utgjør omtrent 10,6 prosent av et standardavvik. Forskjellen mellom MKM- og IV-koeffisienten tyder på at MKM underestimerer effekten av skolestørrelse på matematikkresultater. Dette betyr at jeg ganske sikkert ikke har klart å inkludere nok kontrollvariabler med MKM, og restleddet er korrelert med forklaringsvariabelen.

Tabell 3: IV-estimering med matematikkresultater som avhengig variabel

	Redusert form	Førstesteg	Andresteg
Skolestørrelse	-	-	0,045** (0,008)
Tettsted(3 000-15 000)	7,557** (1,744)	129,952** (8,649)	-
Tettsted(15 000-100 000)	9,945** (1,683)	196,172** (9,230)	-
By(100 000-1 000 000)	9,614** (1,899)	229,108** (10,736)	-
Storby(over 1 000 000)	11,885** (3,699)	234,804** (26,249)	-
F-verdi(førstesteg)	-	161,990	-
R^2	0,278	0,293	0,280
N(Elever)	65074	61543	61543
Cluster (skoler)	2698	2528	2528
Sargan-test (P-verdi)	-	-	0,395

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i matematikk. Alle kontrollvariabler er inkludert. Estimert med instrumentvariabel-metoden. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

6.2.2 Lesing og naturfag

I tabell A6 i appendiks presenteres resultatene fra IV-regresjonen med leseresultater som avhengig variabel. Også her virker det å være en direkte sammenheng mellom instrumentene og prøveresultatene fra redusert form regresjonen. Alle koeffisientene er signifikante på ettprosentnivå, og koeffisientene er også en del større enn i samme regresjon med matematikkresultater. Førstesteget indikerer at instrumentene er relevante. Koeffisientene er alle store og signifikante på ettprosentnivå, og F-verdien er på 162. Heller ikke her kan vi forkaste nullhypotesen om gyldige instrumentvariable i Sargan-testen. Koeffisienten

for skolestørrelse i andresteget er på 0,083, og signifikant på ettprosentnivå. En økning i skolestørrelse med ett standardavvik vil føre til en økning i gjennomsnittresultatet i lesing med 16,9 poeng, noe som utgjør 19,6 prosent av et standardavvik.

I tabell A7 i appendiks presenteres resultatene fra IV-regresjonen med naturfagsresultater som avhengig variabel. Vi ser her samme direkte effekt av skolers lokasjon på resultater i redusert form regresjonen som for matematikk og lesing, og instrumentene virker å være relevante. Koeffisienten for skolestørrelse i andresteget er lik 0,048 og signifikant på ettprosentnivå. En økning i skolestørrelse med ett standardavvik vil føre til en økning i gjennomsnittresultatet i naturfag med 9,8 poeng, noe som utgjør omtrent 11,4 prosent av et standardavvik.

Også resultatene fra lesing og naturfag tyder på at MKM underestimerer effekten av skolestørrelse på elevresultater. IV-resultatene viser at skolestørrelse har en effekt på elevenes resultater. Koeffisientene for skolestørrelse er ikke veldig store, men ved en stor økning i skolestørrelse vil effektene bli store.

6.2.3 Oppsummering

Resultatene fra IV-analysen tyder, i likhet med MKM, på at det er en positiv signifikant effekt av skolestørrelse på elevenes resultater på PISA-undersøkelsen. I hvilken grad vi kan stole på disse resultatene avhenger av hvor gode instrumentene er, som igjen avhenger av i hvilken grad forutsetningene for bruk av IV er oppfylt.

Etter å ha kjørt IV-regresjoner av skolestørrelse på testresultater, virker det ganske klart at instrumentvariablene oppfyller relevanskriteriet. Store og svært signifikante effekter i førstesteget, samt en F-verdi på over 150 i alle regresjonene, gjør at vi kan konkludere med at instrumentene er relevante.

Spørsmålet blir da om vi tror på eksklusjonsforutsetningen. Basert på Sargan-testene som er gjort i analysen virker instrumentene å være gyldige, men dette gir dessverre ikke et fullgodt svar på om forutsetningen er oppfylt. Det at jeg finner en direkte sammenheng mellom instrumentvariablene og testresultater i redusert form ligningen er problematisk, og setter spørsmålsteget ved instrumentenes validitet. Det kan allikevel ikke legges for mye i dette, siden MKM er metoden som benyttes. Noe som er betryggende for analysen er at både MKM og IV-metoden gir en positiv og signifikant effekt av økt skolestørrelse på testresultater. Det er også positivt at effektstørrelsen øker når jeg går fra IV-metoden til MKM, noe som bekrefter mistanken om at MKM underestimerer effekten av skolestørrelse.

Barrow et al. (2013) og Schwartz et al. (2013) benytter seg i likhet med meg av IV-metoden for å håndtere selvseleksjonsproblemet. Hvordan mine resultater ser ut i forhold til deres er derfor av interesse. Barrow et al. (2013) finner en signifikant negativ effekt på

testresultater i matte, naturfag og lesing for elever som går på små skoler med MKM. Ved bruk av IV finner de ingen signifikante effekter på matematikkresultater, i noen tilfeller signifikante negative effekter på leseresultater (avhengig av klassenivå og tester), og det varierer om IV-koeffisientene er større eller mindre enn MKM-koeffisientene. Effektene er i alle tilfeller små, og det konkluderes med at det ikke er noen konsistent effekt av små skoler på testresultater.

I Schwartz et al. (2013) presenteres kun IV-resultater av skolestørrelse på testresultater, siden hovedfokuset i studien er på effekten av skolestørrelse på sannsynligheten for uteksaminering. De finner at både gamle og nye små skoler presterer dårligere enn store skoler i engelsk/lesing, mens det ikke er noen signifikant forskjell på matematikkresultater mellom nye små skoler og store skoler. Den signifikant negative effekten av å gå på en liten skole på testresultater i lesing, er relaterbar til den signifikant positive effekten av skolestørrelse på testresultater jeg finner.

6.3 Heterogene effekter

Jeg vil i dette delkapittelet presentere resultatene fra regresjoner på diverse delutvalg. Dette gjøres for å finne ut om det eksisterer heterogene effekter, og om en begrensning av utvalget vil ha betydning for resultatene fra hovedanalysen.

Jeg vil først presentere resultatene fra regresjoner på enkeltland. Deretter deler jeg opp utvalget avhengig av innvandringsbakgrunn, før jeg ser på om effekten av skolestørrelse er forskjellig avhengig av elevens sosioøkonomiske bakgrunn representert ved antall bøker i hjemmet. Til slutt ser jeg på om effekten av skolestørrelse er forskjellig på tvers av kjønn.

6.3.1 Enkeltland

I tabell 4 presenteres resultatene fra IV-regresjoner gjort på et land av gangen, med matematikkresultater som avhengig variabel. Det er viktig å merke seg at når jeg kjører regresjoner på et land av gangen, reduseres antall observasjoner ganske drastisk i forhold til utgangspunktet, noe som medfører høyere standardfeil.

I alle regresjonene virker instrumentene å være relevante. Selv om F-verdien er mye mindre enn den var for hele datasettet, er den i alle tilfeller langt over 10. Koeffisienten for skolestørrelse er større enn i hovedanalysen for Danmark, Norge og Sverige, og er signifikant på ettprosentnivå i alle de tre tilfellene. I Norge og Sverige er koeffisientene ganske store, og tyder på at disse landene bidrar positivt til skolestørrelseseffekten i hovedanalysen. Koeffisienten for skolestørrelse i Finland skiller seg kraftig fra de andre landene, og den ligger rundt null og er ikke-signifikant.

Effektene endrer seg lite når jeg bruker lese- og naturfagsresultater som avhengig variabel. Resultatene fra disse regresjonene presenteres i tabell A8 og tabell A9 i appendiks. I naturfag er effektene omtrent de samme som i matematikk, og koeffisienten for skolestørrelse er heller ikke her signifikant når jeg bruker det finske utvalget. I lesing er alle koeffisientene for skolestørrelse signifikante, men også her er effekten minst for Finland.

Tabell 4: IV-estimering med matematikkresultater som avhengig variabel, enkeltland

	Danmark	Finland	Norge	Sverige
<i>Andrestep</i>				
Skolestørrelse	0,053** (0,013)	-0,001 (0,012)	0,067** (0,019)	0,082** (0,017)
<i>Førstep</i>				
Tettsted eller by (3 000 til 15 000)	163,495** (19,131)	133,725** (18,697)	125,683** (12,387)	114,033** (18,748)
Tettsted eller by (15 000 til 100 000)	242,849** (19,826)	218,664** (18,420)	161,984** (12,159)	162,368** (19,993)
By (100 000 til 1 000 000)	189,071** (23,429)	256,335** (19,607)	199,138** (16,967)	242,089** (23,700)
Storby (over 1 000 000)	241,297** (28,255)	-	-	274,187** (92,326)
F-verdi (førstep)	41,500	68,960	79,550	31,820
R^2 (førstep)	0,388	0,264	0,287	0,191
R^2 (andresteg)	0,300	0,270	0,198	0,270
N(Elever)	15993	18835	13318	13397
Cluster(Skoler)	744	648	558	578

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i matematikk. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert utenom indikatorvariabler på land. Regresjonene er gjort på utvalg av enkeltland som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

6.3.2 Innvandringsbakgrunn

Tidligere forskning har vist at effektene av skolestørrelse på elevresultater kan ha forskjellig effekt på forskjellige typer elever. Flere studier har funnet at redusert skolestørrelse har en større positiv effekt på minoritets elever og elever med lav sosioøkonomisk bakgrunn (Cotton, 1996; Lee og Smith, 1997). For å undersøke om en slik forskjell er til stede i mitt utvalg, vil jeg kjøre regresjoner på utvalg av elever med ulike innvandringsbakgrunn.

I tabell 5 presenteres resultatene fra IV-regresjon av skolestørrelse på matematikkresultater hvor utvalget er delt opp i innvandrere (første- og andregenerasjons) og etnisk nordiske

elever. Koeffisienten for skolestørrelse er på 0,009 og ikke signifikant i innvandringsutvalget.

Ved å kjøre samme regresjon på utvalget av etnisk nordiske elever, får jeg en koeffisient på 0,05 som er signifikant på ettprosentnivå. F-verdien i førstesteget er nå på 163,9, og tyder på relevante instrumenter.

Når jeg bruker naturfagsresultater som avhengig variabel får jeg omtrent samme forskjell mellom innvandrere og etnisk nordiske elever som i matematikk. I lesing virker det derimot ikke å være noen forskjell i koeffisienten for skolestørrelse mellom de to utvalgene. Resultatene fra disse regresjonene finnes i tabell A10 og A11 i appendiks.

Tabell 5: IV-estimering med matematikkresultater som avhengig variabel, innvandringsbakgrunn

	Innvandrere	Etnisk nordiske
Skolestørrelse	0,009 (0,026)	0,050** (0,007)
F-verdi (førstesteg)	44,670	163,910
R^2 (førstesteg)	0,162	0,309
R^2 (andresteg)	0,190	0,210
N(Elever)	7150	53303
Cluster(Skoler)	1657	2512

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i matematikk. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på innvandringsbakgrunn. Regresjoner er kjørt på utvalg avhengig av innvandringsbakgrunn som indikert i kolonneoverskrifter. Standardfeil i parenteser er korrigert for å ta høyde for clustering innen skoler.

6.3.3 Sosioøkonomisk bakgrunn

Effekten av skolestørrelse på testresultater kan også være forskjellig avhengig av elevenes sosioøkonomiske bakgrunn. Siden jeg i min analyse bruker antall bøker i hjemmet som en proxy på elevenes sosioøkonomiske bakgrunn, vil jeg se på om jeg oppnår forskjellige resultater for elever med forskjellig antall bøker i hjemmet.

I tabell 6 presenteres resultatene fra IV-regresjon av skolestørrelse på matematikkresultater hvor utvalget er delt opp i elever med mellom 0 og 25 bøker i hjemmet (lav sosioøkonomisk bakgrunn), elever med mellom 26 og 200 bøker i hjemmet (middels sosioøkonomisk bakgrunn), og elever med over 200 bøker i hjemmet (høy sosioøkonomisk bakgrunn). IV-regresjonen på utvalget av elever med få bøker i hjemmet gir en koeffisient på skolestørrelse på 0,027 som er signifikant på femprosentnivå. Koeffisienten er større og mer signifikant

i de to andre utvalgene, og størst i utvalget av elever med over 200 bøker i hjemmet, hvor koeffisienten er mer enn dobbelt så stor som koeffisienten på utvalget med få bøker i hjemmet.

Når jeg bruker lese- og naturfagsresultater som avhengig variabel, ser jeg nøyaktig den samme sammenhengen som når jeg bruker resultater i matematikk. Det henvises til tabell A12 og A13 i appendiks for resultater fra disse regresjonene.

Tabell 6: IV-estimering med matematikkresultater som avhengig variabel, sosioøkonomisk bakgrunn (SES)

	Lav SES	Middels SES	Høy SES
Skolestørrelse	0,027*	0,042**	0,065**
	(0,012)	(0,008)	(0,011)
F-verdi (førstesteg)	109,350	157,280	136,640
R^2 (førstesteg)	0,292	0,296	0,288
R^2 (andresteg)	0,192	0,142	0,124
N(Elever)	13302	30972	15809
Cluster(Skoler)	2378	2502	2427

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i matematikk. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på antall bøker i hjemmet. Regresjoner er kjørt på utvalg avhengig av sosioøkonomisk bakgrunn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

6.3.4 Kjønn

I litteraturen er det foreslått at mindre skoler fører til mindre vold og atferdsproblemer grunnet mindre anonymitet (Meier, 1996). Siden problemer som dette ofte er mer utbredt blant gutter enn blant jenter, vil jeg se på om effekten av skolestørrelse er forskjellig avhengig av kjønn.

I tabell 7 presenteres resultatene fra IV-regresjon av skolestørrelse på matematikkresultater på tvers av kjønn. Siden det er omtrent like mange gutter som jenter, betyr dette at utvalget deles i to. IV-regresjonen på utvalget som kun består av gutter gir en koeffisient på skolestørrelse på 0,053 som er signifikant på ettprosentnivå. Samme regresjonen på utvalget av jenter gir en koeffisient på 0,039 som også er signifikant på ettprosentnivå.

Også med lese- og naturfagsresultater som avhengig variabel virker effekten av skolestørrelse på testresultater å være noe større for gutter enn for jenter (se tabell A14 og A15 i appendiks). Resultatene tyder på at gutter har en større positiv effekt av økt skolestørrelse enn jenter.

Tabell 7: IV-estimering med matematikkresultater som avhengig variabel, kjønn

	Gutter	Jenter
Skolestørrelse	0,051** (0,009)	0,037** (0,008)
F-verdi (førstesteg)	154,430	157,460
R^2 (førstesteg)	0,294	0,294
R^2 (andresteg)	0,266	0,289
N(Elever)	30775	30768
Cluster(Skoler)	2499	2497

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i matematikk. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabel på kjønn. Regresjoner er kjørt på utvalg avhengig av kjønn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

6.3.5 Oppsummering

Jeg har i dette delkapittelet sett på om det eksisterer heterogene effekter av skolestørrelse på testresultater i de tre fagene. Analysen gjort på enkeltland tyder på at det er en større effekt av økt skolestørrelse i Danmark, Norge og Sverige enn i Finland, hvor sammenhengen virker å være svak. Ved bruk av MKM finner jeg en signifikant effekt av skolestørrelse på matematikkresultater i Finland, mens i de andre landene er effektene også signifikante, men mindre enn med IV-metoden. Dette tyder på at retningen på skjevheten ved MKM går motsatt vei i Finland i forhold til de andre landene. Hadde gjennomsnittsstørrelsen på finske skoler i utvalget skilt seg ut fra gjennomsnittsstørrelsen på skolene i de andre landene, kunne dette vært en mulig forklaring på forskjellen, men den deskriptive statistikken viser at Finland har skoler i samme størrelsesorden som Danmark og Sverige. Det må derfor være andre grunner til den ulike effekten. Finland presterer mye bedre enn sine nordiske naboland i alle fag, så dette er en mulig forklaring. Hvis en antar at flinke elever er mer tjent med økt skolestørrelse enn mindre flinke elever, burde vi se en større effekt av skolestørrelse på testresultater i Finland enn i de andre landene, så derfor tror jeg heller ikke dette kan forklare forskjellene. Kanskje er det heller et spørsmål om lærerkvalitet og seleksjon blant lærere. Jeg ser for meg at det kan være slik at mindre dyktige lærere kan bli flinkere ved økt skolestørrelse fordi de får flere kollegaer og kanskje økt sannsynlighet for at det er dyktige lærere på skolen som de kan dra nytte av. Hvis lærerkvaliteten er høyere i Finland enn i resten av Norden, fordi lærerstudiet i Finland er mer krevende og tiltrekker seg mange av de beste studentene (Sahlberg, 2013), vil lærerne her være mer

selvgående og mindre avhengige av andre (både på små og store skoler). Dette vil gi en forskjell i skolestørrelseseffekter og er en mulig forklaring på forskjellen mellom Finland og de andre landene, men det er dessverre ingen måte å teste for dette med de data jeg har.

Resultatene fra analysen gjort på elever med forskjellig innvandringsbakgrunn indikerer at etnisk nordiske elever er bedre tjent med økt skolestørrelse enn minoritets elever (som ikke påvirkes signifikant), i hvert fall i realfag. Resultatene tyder ikke på at minoritets elever er tjent med mindre skolestørrelse som vist i tidligere forskning (Cotton, 1996), men forskjellen mellom minoritets elever og etnisk nordiske elever er allikevel relaterbar til denne forskningen ved at de etnisk nordiske elevene er bedre tjent med økt skolestørrelse.

Det virker å være en klar forskjell i effektstørrelse avhengig av elevenes sosioøkonomiske bakgrunn. Den positive effekten av økt skolestørrelse er større for elever med høy sosioøkonomisk bakgrunn enn for elever med lav sosioøkonomisk bakgrunn. Resultatene bør dog tolkes med varsomhet, fordi det er uklart hvor godt antall bøker i hjemmet fungerer som en proxy på elevenes sosioøkonomiske bakgrunn.

Analysen gjort på utvalg av forskjellige kjønn indikerer at gutter har en større positiv effekt av økt skolestørrelse enn jenter, men forskjellen i koeffisienter er ikke veldig stor i noen av fagene. Dette er det motsatte av hva jeg forventet å finne, så resultatene må sies å være overraskende.

En av svakhetene med å se på heterogene effekter på den måten jeg har gjort, er at antall observasjoner faller ganske drastisk i en del av regresjonene. Dette fører til mindre presis estimering enn i hovedanalysen.

7 Oppsummering og diskusjon

I denne oppgaven har jeg sett på hvordan skolestørrelse påvirker elevers prestasjoner på PISA-undersøkelsene i de nordiske landene. Oppgaven baserer seg hovedsakelig på studier gjort i USA, hvor funnene har vært motstridende. Målet har vært å undersøke i hvilken grad resultatene fra disse studiene er overførbare til Norden, hvor skolesystemet er ganske annerledes og skolene er mye mindre. Hypotesen min har vært at økt skolestørrelse vil føre til bedre elevresultater på grunn av stordriftsfordeler.

Jeg starter analysen med vanlig minste kvadraters metode (MKM) på en enkel skoleproduktfunksjon, hvor jeg på best mulig måte prøver å inkludere kontrollvariabler som potensielt kan ha en påvirkning på elevenes resultater. Jeg finner at skolestørrelse har en signifikant positiv effekt på elevenes testresultater i både matematikk, lesing og naturfag, men effektstørrelsene er ikke store. For å se om det eksisterer en ikke-lineær sammenheng inkluderer jeg et kvadratisk ledd for skolestørrelse, men jeg finner ingen signifikant effekt.

På grunn av svakhetene ved bruk av MKM, og en mistanke om skjeve resultater på grunn av seleksjon og utelatte variabler, gjennomfører jeg så en instrumentvariabel-analyse (IV). Som instrumentvariabler benytter jeg indikatorer på lokalmiljøets størrelse hvor skolene ligger, motivert av tidligere forskning. Også ved bruk av IV finner jeg en signifikant positiv effekt av skolestørrelse på elevenes resultater i de tre testfagene, og koeffisientene mer enn dobler seg i forhold til MKM. Selv om det er mistanke om utelatte variable ved MKM, og det kan stilles spørsmål ved eksklusjonsforutsetningen ved IV-metoden, er det interessant at begge metodene gir en positiv signifikant effekt.

Ved å dele opp utvalget i undergrupper og repetere IV-regresjonene på delutvalgene, finner jeg ut om det eksisterer heterogene effekter, og om noen undergrupper driver resultatene i hovedanalysen. Jeg finner at effektene av skolestørrelse på testresultater er særlig stor i Norge og Sverige, mens det i Finland virker å være en svakere sammenheng. Videre finner jeg at etnisk nordiske elever er bedre tjent med økt skolestørrelse enn elever med innvandringsbakgrunn, og at den positive effekten av økt skolestørrelse er større for elever med høy sosioøkonomisk bakgrunn enn for elever med lav sosioøkonomisk bakgrunn. Gutter virker også å ha en noe større positiv effekt av økt skolestørrelse enn hva jenter har, men forskjellen er liten.

Resultatene fra min analyse skiller seg en del fra funnene gjort i nyere amerikanske studier, som finner ingen eller små effekter av skolestørrelse på testresultater. Det kan være flere grunner til at resultatene mine skiller seg fra deres, og jeg ser spesielt for meg tre viktige årsaker. For det første inkluderer mitt datasett skoler som er langt mindre i størrelse enn de amerikanske. For det andre benyttes forskjellige resultatmålmål. PISA-prøvene tar sikte på å måle elevenes evne til å bruke kunnskaper og erfaringer i aktuelle situasjoner,

mens de amerikanske prøvene som regel er mer pensumbaserte. For det tredje mistenker jeg at datasettene i de amerikanske studiene består av en langt større andel elever med ufordelaktig bakgrunn sammenlignet med mitt datasett. Det er blitt vist i tidligere studier (Page et al., 2002), og også i min studie, at elever med høy sosioøkonomisk bakgrunn er tjent med større skoler, og derfor kan en slik forskjell i data føre til forskjellige effekter av skolestørrelse på testresultater.

Selv om resultatene fra min analyse må tolkes med forsiktighet på grunn av usikkerheten rundt metodene som brukes, gir de allikevel en indikasjon på at økt skolestørrelse kan gi bedre testresultater. Sammen med funnet i Falch et al. (2008) om skalafordeler i norske skoler, kan dette bidra som et argument for skolesammenslåinger i Norge. Hvis skolene som deltar i PISA-undersøkelsene er et representativt utvalg, noe det skal være, viser den deskriptive statistikken at de norske skolene i gjennomsnitt er ganske mye mindre enn sammenlignbare skoler i Danmark, Finland og Sverige. Dette betyr at det er et potensiale for å øke størrelsen på en del skoler i Norge, som kan skje ved nedleggelse av små skoler samtidig som det enten bygges nye skoler, eller kapasiteten på eksisterende skoler økes. Sammenslåinger av skoler er også et alternativ. Siden Norge er et spredtbygd land, blir det da et spørsmål om dette er gjennomførbart uten å øke reiseavstanden for en del elever for mye.

Det er også en del andre hensyn som må tas i forbindelse med endringer i skolestørrelse. Mindre skolestørrelse er tidligere blitt vist å ha en positiv innvirkning på en rekke andre elevutfall enn testresultater, og disse effektene kan ikke overses. Fordi testresultater er blitt sagt å være et ufullstendig mål på skolekvalitet, motiverer dette for videre forskning på området i Norden med andre resultatmål. Mer komplette datasett, som følger elever over flere år, kan gi informasjon om skolestørrelses påvirkning på elevutfall senere i livet, som for eksempel suksess i arbeidslivet. Også en grundigere analyse hvor andre funksjonsformer benyttes kan være å anbefale for fremtidige undersøkelser, særlig for å se nærmere på om effekten av skolestørrelse kan være ikke-lineær. Dette er viktig for å kunne komme nærmere en løsning på hvordan vi kan forbedre utdanningssektoren i Norge.

Referanser

- Barker, R. og Gump, P. (1964): "Big School, Small School." Stanford, CA: Stanford University Press.
- Barrow, L., Claessens, A. og Schanzenbach, D. W. (2013): "The Impact of Chicago's Small High School Initiative." March, Working Paper.
- Begg, D., Fischer, S. og Dornbusch, R. (2008): *Economics (ninth edition)*. McGraw-Hill, USA.
- Bound, J., Jaeger, D. A. og Baker, R. M. (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90, 443–450.
- Buzacott, J. A. (1982): *Scale in Production Systems*. Pergamon.
- Conant, J. B. (1959): *The American High School Today*. McGraw-Hill.
- Cotton, K. (1996): *School Size, School Climate, and Student Performance*. Northwest Regional Educational Laboratory.
- Dagbladet (2013): "355 grunnskoler er nedlagt siden 2005." (<http://www.dagbladet.no/2013/08/02/nyheter/politikk/utdanning/innenriks/valg13/28509349/>).
- Edington, E. D. og Gardener, C. E. (1984): "The Relationship of School Size to Scores in the Affective Domain From the Montana Testing Service Examination." *Education* 105, 40–45.
- Falch, T., Rønning, M. og Strøm, B. (2008): "A Cost Model of Schools: School Size, School Structure and Student Composition." I N.C. Soguel og P. Jaccard, Governance and Performance of Education Systems, (s. 247-265).
- Fowler Jr., W. J. og Walberg, H. J. (1991): "School Size, Characteristics, and Outcomes." *Educational Evaluation and Policy Analysis* 13, 189–202.
- Fox, W. F. (1981): "Reviewing Economies of Size in Education." *Journal of Education Finance* 6, 273–296.
- Friedkin, N. E. og Necochea, J. (1988): "School System Size and Performance: A Contingency Perspective." *Educational Evaluation and Policy Analysis* 10, 237–249.
- Gooding, R. Z. og Wagner, J. A. (1985): "A Meta-analytic Review of the Relationship Between Size and Performance: The Productivity and Efficiency of Organizations and Their Subunits." *Administrative Science Quarterly* 30, 462–481.

- Haller, E. J. (1992): “High School Size and Student Indiscipline: Another Aspect of the School Consolidation Issue?” *Educational Evaluation and Policy Analysis* 14, 145–156.
- Hanushek, E. A. (2003): “The Failure of Input-based Schooling Policies.” *Economic Journal* 113, 64–98.
- Hanushek, E. A., Link, S. og Woessmann, L. (2013): “Does School Autonomy Make Sense Everywhere? Panel Estimates From PISA.” *Journal of Development Economics* 104, 212–232.
- Heinesen, E. (2004): “School District Size and Student Educational Attainment: Evidence From Denmark.” *Economics of Education Review* 24, 677–689.
- Howley, C. (1997): “Dumbing Down by Sizing Up.” *The School Administrator* 54, 24–30.
- Jewell, R. W. (1989): “School and School District Size Relationships: Costs, Results, Minorities, and Private School Enrollments.” *Education and Urban Society* 21, 140–153.
- Kenny, L. W. (1982): “Economies of Scale in Schooling.” *Economics of Education Review* 2, 1–24.
- Kjærnsli, M. og Olsen, R. V. (2013): *Fortsatt en vei å gå*. Universitetsforlaget.
- Krueger, A. (2000): “An Economist’s View of Class Size Research.”
- Kuziemko, I. (2004): “Using Shocks to School Enrollment to Estimate the Effect of School Size on Student Achievement.” *Economics of Education Review* 25, 63–75.
- Lee, V. E., Bryk, A. S. og Smith, J. B. (1993): “The Organization of Effective High Schools.” *Review of research in education* 19, 171–267.
- Lee, V. E. og Smith, J. B. (1997): “High School Size: Which Works Best and for Whom?” *Educational Evaluation and Policy Analysis* 19, 205–227.
- McGuffey, C. og Brown, C. L. (1978): “The Relationship of School Size and Rate of School Plant Utilization to Cost Variations of Maintenance and Operation.” *American Educational Research Journal* 15, 373–378.
- Meier, D. W. (1996): “The Big Benefits of Smallness.” *Educational Leadership* 54, 12–15.
- Monk, D. (1987): “Secondary School Size and Curriculum Comprehensiveness.” *Economics of Education Review* 6, 137–150.
- Murray, M. P. (2006): “Avoiding Invalid Instruments and Coping with Weak Instruments.” *Journal of Economic Perspectives* 20, 111–132.
- OECD (2013): “Education at a Glance 2013: OECD Indicators.”

- Page, L., Layzer, C., Schimmenti, J., Bernstein, L. og Horst, L. (2002): “National Evaluation of Smaller Learning Communities: Literature Review.”
- Pittman, R. B. og Haugwout, P. (1987): “Influence of High School Size on Dropout Rate.” *Educational Evaluation and Policy Analysis* 9, 337–343.
- Sahlberg, P. (2013): “Teachers as Leaders in Finland.” *Educational Leadership* 71, 36–40.
- Schneider, B., Wyse, A. E. og Keesler, V. (2007): “Is Small Really Better? Testing Some Assumptions about High School Size.” *Brookings Papers on Education Policy* 9, 15–47.
- Schwartz, A. E., Stiefel, L. og Wiswall, M. (2013): “Do Small Schools Improve Performance in Large, Urban Districts? Causal Evidence From New York City.” *Journal of Urban Economics* 77, 27–40.
- Smith, D. og DeYoung, A. (1988): “Big School vs. Small School: Conceptual, Empirical, and Political Perspectives on the Reemerging Debate.” *Journal of Rural and Small Schools* 2, 2–11.
- Sollien, T. H. (2008): “Sammenheng mellom skolestørrelse og kvalitet.” Asplan Viak.
- Stiefel, L., Berne, R., Iatarola, P. og Fruchter, N. (2000): “High School Size: Effects on Budgets and Performance in New York City.” *Educational Evaluation and Policy Analysis* 22, 27–39.
- Strang, D. (1987): “The Administrative Transformation of American Education: School District Consolidation, 1938-1980.” *Administrative Science Quarterly* 3, 352–366.
- Taylor, J. og Bradley, S. (2000): “Resource Utilization and Economies of Size in Secondary Schools.” *Bulletin of Economic Research* 52, 123–150.
- Todd, P. E. og Wolpin, K. I. (2003): “On the Specification and Estimation of the Production Function for Cognitive Achievement.” *The Economic Journal* 113, F3–F33.
- Walberg, H. J. og Walberg, I., H. J. (1994): “Losing Local Control.” *Educational Researcher* 23, 19–26.
- Woolridge, J. M. (2009): *Introductory Econometrics: A Modern Approach (fourth edition)*. South-Western Cengage Learning, Canada.

A Appendiks

A.1 Deskriptiv statistikk

Tabell A1: Deskriptiv statistikk

Variabler	Gj.snitt	Std.	Min.	Maks.	Missing(i %)
PISA-matematikkresultat	502,856	86,334	85,130	793,885	0
PISA-leseresultat	502,513	92,581	54,664	831,459	0
PISA-naturfagsresultat	506,382	95,010	135,433	829,112	0
Elev- og familiekarakteristikk					
Jente	0,499	0,500	0	1	0
Alder	15,742	0,287	15,170	16,330	0,002
<i>Immigrasjonsbakgrunn</i>					
Etnisk nordisk	0,880	0,325	0	1	1,778
Førstegenerasjonsinnvandrere	0,072	0,259	0	1	1,778
Andregenerasjonsinnvandrere	0,047	0,212	0	1	1,778
<i>Språk i hjemmet</i>					
Samme som testspråket	0,906	0,291	0	1	4,254
Annen nasjonal dialekt	0,081	0,272	0	1	4,254
Annet språk	0,013	0,114	0	1	4,254
<i>Mors utdanning</i>					
Ingen	0,014	0,116	0	1	4,841
Grunnskoleutdanning	0,019	0,135	0	1	4,841
Ungdomskoleutdanning	0,074	0,262	0	1	4,841
Videregående utdanning	0,084	0,277	0	1	4,841
Utdanning etter vgs annet enn univ.	0,197	0,398	0	1	4,841
Kortvarig universitetsutdanning	0,331	0,470	0	1	4,841
Bachelor eller bedre	0,282	0,450	0	1	4,841
<i>Fars utdanning</i>					
Ingen	0,014	0,116	0	1	7,467
Grunnskoleutdanning	0,025	0,156	0	1	7,467

Fortsetter på neste side

TabellA1 – fortsetter fra forrige side

Variabler	Gj.snitt	Std.	Min.	Maks.	Missing(i %)
Ungdomskoleutdanning	0,104	0,306	0	1	7,467
Videregående utdanning	0,133	0,340	0	1	7,467
Utdanning etter vgs annet enn univ.	0,201	0,401	0	1	7,467
Kortvarig universitetsutdannelse	0,253	0,435	0	1	7,467
Bachelor eller bedre	0,269	0,443	0	1	7,467
<i>Bøker i hjemmet</i>					
0-10	0,096	0,294	0	1	2,422
11-25	0,126	0,332	0	1	2,422
26-100	0,312	0,463	0	1	2,422
101-200	0,203	0,402	0	1	2,422
201-500	0,174	0,379	0	1	2,422
Over 500	0,089	0,285	0	1	2,422
Skolekarakteristikk					
Skolestørrelse	424,608	203,126	3	2512	5,426
Privatskole	0,092	0,289	0	1	0
Andel offentlig finansiering	0,980	0,082	0	1	0,047
Andel fullt sertifiserte lærere	0,896	0,178	0	1	0,255
<i>Mangel på mattelærere</i>					
Ikke i det hele tatt	0,711	0,453	0	1	3,894
Veldig lite	0,208	0,406	0	1	3,894
Til en viss grad	0,077	0,267	0	1	3,894
I stor grad	0,004	0,067	0	1	3,894
<i>Mangel på lærere i lesing</i>					
Ikke i det hele tatt	0,771	0,420	0	1	3,806
Veldig lite	0,183	0,387	0	1	3,806
Til en viss grad	0,043	0,203	0	1	3,806
I stor grad	0,003	0,052	0	1	3,806
<i>Mangel på naturfagslærere</i>					

Fortsetter på neste side

TabellA1 – fortsetter fra forrige side

Variabler	Gj.snitt	Std.	Min.	Maks.	Missing(i %)
Ikke i det hele tatt	0,668	0,471	0	1	3,885
Veldig lite	0,217	0,412	0	1	3,885
Til en viss grad	0,108	0,311	0	1	3,885
I stor grad	0,007	0,081	0	1	3,885
<i>Skolens beliggenhet</i>					
Lite sted (under 3 000)	0,157	0,364	0	1	3,017
Tettsted eller by (3 000 til 15 000)	0,268	0,443	0	1	3,017
Tettsted eller by (15 000 til 100 000)	0,329	0,470	0	1	3,017
By (100 000 til 1 000 000)	0,222	0,416	0	1	3,017
Storby (over 1 000 000)	0,024	0,153	0	1	3,017
Andre kontroller					
<i>Dummy for land</i>					
Danmark	0,276	0,447	0	1	0
Finland	0,297	0,457	0	1	0
Norge	0,216	0,411	0	1	0
Sverige	0,211	0,408	0	1	0
<i>Dummy for år</i>					
2006	0,283	0,450	0	1	0
2009	0,322	0,467	0	1	0
2012	0,395	0,489	0	1	0

Datasettet består av PISA 2006, 2009 og 2012 med observasjoner for Danmark, Finland, Norge og Sverige.

Tabell A2: Deskriptiv statistikk, skolestørrelse

Variabler	Gj.snitt	Standardavvik	Min.	Maks.	Missing(i %)
2006					
Skolestørrelse i Danmark	459,760	218,269	48	906	21,690
Skolestørrelse i Finland	424,143	169,885	32	924	1,888
Skolestørrelse i Norge	321,418	122,596	16	597	2,003
Skolestørrelse i Sverige	485,181	195,702	5	1426	1,733
2009					
Skolestørrelse i Danmark	491,485	213,413	70	1467	1,435
Skolestørrelse i Finland	410,314	187,175	12	1967	1,91
Skolestørrelse i Norge	343,540	137,041	27	720	6,438
Skolestørrelse i Sverige	429,881	219,429	3	1995	5,956
2012					
Skolestørrelse i Danmark	505,282	247,047	12	1237	11,710
Skolestørrelse i Finland	428,065	188,442	10	946	3,602
Skolestørrelse i Norge	342,597	135,508	18	750	6,957
Skolestørrelse i Sverige	404,438	234,908	14	2512	0

A.2 Regresjoner

Tabell A3: MKM-estimering med leseresultater som avhengig variabel

	(1)	(2)	(3)	(4)
Skolestørrelse	0,026** (0,005)	0,032** (0,003)	0,023** (0,004)	0,027** (0,004)
Kontroller for elevkarakteristikk	Nei	Ja	Ja	Ja
Kontroller for skolekarakteristikk	Nei	Nei	Ja	Ja
Dummyer for land og år	Nei	Nei	Nei	Ja
R^2	0,003	0,293	0,300	0,314
N(Elever)	61543	61543	61543	61543
Cluster(Skoler)	2528	2528	2528	2528

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i lesing. Estimert med minste kvadraters metode. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i parenteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A4: MKM-estimering med naturfagsresultater som avhengig variabel

	(1)	(2)	(3)	(4)
Skolestørrelse	0,018** (0,005)	0,025** (0,004)	0,021** (0,004)	0,022** (0,004)
Kontroller for elevkarakteristikk	Nei	Ja	Ja	Ja
Kontroller for skolekarakteristikk	Nei	Nei	Ja	Ja
Dummyer for land og år	Nei	Nei	Nei	Ja
R^2	0,001	0,264	0,273	0,312
N(Elever)	61543	61543	61543	61543
Cluster(Skoler)	2528	2528	2528	2528

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i naturfag. Estimert med minste kvadraters metode. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i parenteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A5: MKM-estimering med alle kontrollvariabler inkludert. Avhengige variabler er testresultater i de tre fagene.

Variabler	Matte	Lesing	Naturfag
Skolestørrelse	0,022** (0,003)	0,027** (0,004)	0,022** (0,004)
Elev- og familiekarakteristikk			
Jente	-11,603** (0,650)	38,429** (0,695)	-4,264** (0,711)
Alder	10,549** (1,070)	11,303** (1,124)	10,882** (1,145)
MissingAlder	214,781** (16,957)	128,973** (18,140)	173,995** (18,161)
<i>Immigrasjonsbakgrunn</i>			
Etnisk nordisk			
Førstegenerasjonsinnvandrere	-35,820** (2,079)	-35,439** (2,300)	-40,999** (2,253)
Andregenerasjonsinnvandrere	-22,901** (1,743)	-13,986** (1,889)	-27,105** (1,837)
MissingImmigrasjonsbakgrunn	-30,156** (3,072)	-33,697** (3,364)	-39,765** (3,286)
<i>Språk i hjemmet</i>			
Samme som testspråket			
Annen nasjonal dialekt	-15,784** (1,741)	-27,476** (1,945)	-28,268** (1,918)
Annet språk	-12,599** (3,593)	-12,818** (3,810)	-15,642** (3,691)
MissingSpråk	-40,475** (1,950)	-42,704** (2,029)	-42,823** (2,094)
<i>Mors utdanning</i>			
Ingen			
Grunnskoleutdanning	5,280 (3,734)	11,046** (4,178)	8,469* (3,996)
Ungdomskoleutdanning	7,553** (3,104)	13,371** (3,671)	12,876** (3,354)
Videregående utdanning	15,652** (3,125)	20,016** (3,678)	22,589** (3,395)

Fortsetter på neste side

TabellA5 – fortsetter fra forrige side

Variabler	Matte	Lesing	Naturfag
Utdanning etter vgs annet enn universitet	17,020** (3,025)	23,794** (3,586)	24,737** (3,253)
Kortvarig universitetsutdannelse	26,404** (2,991)	32,210** (3,573)	34,305** (3,234)
Bachelor eller bedre	21,130** (3,045)	27,509** (3,621)	29,503** (3,291)
MissingMorsutdannelse	-6,188† (3,334)	-7,719* (3,901)	-4,899 (3,614)
<i>Fars utdannelse</i>			
Ingen			
Grunnskoleutdanning	18,999** (3,537)	25,761** (3,846)	23,248** (3,778)
Ungdomskoleutdanning	19,964** (3,061)	26,895** (3,370)	21,421** (3,135)
Videregående utdanning	31,296** (3,056)	36,379** (3,369)	31,693** (3,151)
Utdanning etter vgs annet enn universitet	27,856** (3,014)	32,816** (3,302)	27,102** (3,086)
Kortvarig universitetsutdannelse	32,349** (2,993)	39,059** (3,301)	34,706** (3,077)
Bachelor eller bedre	36,599** (3,002)	41,579** (3,312)	37,621** (3,091)
MissingFarsutdannelse	6,282* (3,084)	9,425** (3,458)	5,287 (3,250)
<i>Bøker i hjemmet</i>			
0-10			
11-25	12,775** (1,290)	13,757** (1,364)	16,958** (1,380)
26-100	34,626** (1,153)	37,402** (1,245)	39,273** (1,233)
101-200	55,808** (1,269)	59,871** (1,356)	63,723** (1,346)
201-500	77,282** (1,337)	77,525** (1,439)	83,011** (1,443)
Over 500	84,032** (1,635)	79,604** (1,733)	88,772** (1,735)

Fortsetter på neste side

TabellA5 – fortsetter fra forrige side

Variabler	Matte	Lesing	Naturfag
MissingBøker	32,352** (2,704)	20,469** (3,047)	29,560** (2,931)
Skolekarakteristikk			
Privatskole	8,302** (2,692)	14,525** (2,957)	12,798** (3,124)
Andel offentlig finansiering	-0,138 (0,096)	0,016 (0,110)	-0,044 (0,108)
MissingOffentligfinansiering	-9,523 (10,456)	3,630 (11,688)	-3,080 (11,582)
Andel fullt sertifiserte lærere	-0,262 (3,404)	2,840 (4,341)	0,898 (3,810)
MissingSertifisertelærere	2,838 (3,668)	4,484 (4,445)	1,355 (4,006)
<i>Mangel på lærere i det aktuelle fag</i>			
Ikke i det hele tatt			
Veldig lite	-3,280* (1,484)	-4,498** (1,737)	-6,063** (1,569)
Til en viss grad	-6,400** (2,036)	-3,493 (3,226)	-4,411* (1,927)
I stor grad	1,177 (4,873)	19,342† (10,092)	-0,667 (6,307)
MissingLærermangel	10,802† (6,110)	5,195 (8,465)	10,674 (5,180)
<i>Skolens beliggenhet</i>			
Lite sted (under 3000)			
Tettsted eller by (3000 til 15 000)	4,655** (1,810)	7,759** (2,064)	5,377** (1,963)
Tettsted eller by (15 000 til 100 000)	5,860** (1,816)	11,855** (2,039)	6,163** (2,000)
By (100 000 til 1 000 000)	4,468* (2,059)	11,877** (2,267)	4,950* (2,209)
Storby (over 1 000 000)	6,875† (3,614)	18,052** (4,015)	9,526** (4,051)
MissingSkolensbeliggenhet	13,884 (9,344)	40,094** (9,773)	15,970 (13,019)
Fortsetter på neste side			

TabellA5 – fortsetter fra forrige side

Variabler	Matte	Lesing	Naturfag
Andre kontroller			
<i>Dummy for land</i>			
Danmark	12,599** (2,161)	-3,372 (2,292)	5,738** (2,197)
Finland	35,720** (1,680)	26,107** (2,019)	48,687** (1,823)
Sverige	0,900 (1,829)	-1,609 (2,203)	2,216 (1,938)
<i>Dummy for år</i>			
2009	-4,632** (1,492)	-1,419 (1,645)	-0,989 (1,506)
2012	-13,043** (1,373)	-2,806 [†] (1,654)	-2,962* (1,485)
R^2	0,278	0,314	0,312
N(Elever)	61543	61543	61543
Cluster(Skoler)	2528	2528	2528

Avhengige variabler er testresultater i de tre fagene. Estimert med minste kvadraters metode. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i parenteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A6: IV-estimering med leseresultater som avhengig variabel

	Redusert form	Førstesteg	Andresteg
Skolestørrelse	-	-	0,083** (0,009)
Tettsted(3 000-15 000)	11,490** (1,989)	130,826** (8,606)	-
Tettsted(15 000-100 000)	17,112** (1,868)	196,684** (9,238)	-
By(100 000-1 000 000)	18,567** (2,099)	230,408** (10,792)	-
Storby(over 1 000 000)	24,591** (4,040)	234,075** (26,425)	-
F-verdi	-	162,020	-
R^2	0,313	0,292	0,303
N(Elever)	65074	61543	61543
Cluster (skoler)	2698	2528	2528
Sargan-test (P-verdi)	-	-	0,451

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i lesing. Alle kontrollvariabler er inkludert. Estimert med instrumentvariabel-metoden. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i parenteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A7: IV-estimering med naturfagsresultater som avhengig variabel

	Redusert form	Førstesteg	Andrestep
Skolestørrelse	-	-	0,048** (0,008)
Tettsted(3 000-15 000)	8,676** (1,881)	130,722** (8,591)	-
Tettsted(15 000-100 000)	10,832** (1,836)	196,149** (9,270)	-
By(100 000-1 000 000)	10,623** (2,056)	229,475** (10,761)	-
Storby(over 1 000 000)	15,062** (4,059)	234,534** (26,262)	-
F-verdi	-	161,060	-
R^2	0,312	0,294	0,310
N(Elever)	65074	61543	61543
Cluster (skoler)	2698	2528	2528
Sargan-test (P-verdi)	-	-	0,282

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i naturfag. Alle kontrollvariabler er inkludert. Estimert med instrumentvariabel-metoden. Utvalget består av Danmark, Finland, Norge og Sverige for årene 2006, 2009 og 2012. Standardfeil i parenteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A8: IV-estimering med leseresultater som avhengig variabel, enkeltland

	Danmark	Finland	Norge	Sverige
Skolestørrelse	0,077** (0,013)	0,058** (0,014)	0,101** (0,021)	0,094** (0,019)
F-verdi (førstesteg)	40,850	71,590	81,140	31,940
R^2 (førstesteg)	0,386	0,262	0,293	0,181
R^2 (andresteg)	0,291	0,330	0,272	0,305
N(Elever)	15993	18835	13318	13397
Cluster(Skoler)	744	648	558	578

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i lesing. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på land. Regresjonene er gjort på utvalg av enkeltland som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A9: IV på naturfagsresultater, enkeltland

	Danmark	Finland	Norge	Sverige
Skolestørrelse	0,055** (0,013)	0,010 (0,013)	0,066** (0,021)	0,073** (0,018)
F-verdi (førstesteg)	41,97	65,980	77,200	31,910
R^2 (førstesteg)	0,386	0,267	0,284	0,183
R^2 (andresteg)	0,314	0,267	0,226	0,282
N(Elever)	15993	18835	13318	13397
Cluster(Skoler)	744	648	558	578

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i naturfag. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på land. Regresjonene er gjort på utvalg av enkeltland som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A10: IV-estimering med leseresultater som avhengig variabel, innvandringsbakgrunn

	Innvandrere	Etnisk nordiske
Skolestørrelse	0,088** (0,029)	0,086** (0,008)
F-verdi (førstesteg)	45,020	164,200
R^2 (førstesteg)	0,162	0,308
R^2 (andresteg)	0,213	0,241
N(Elever)	7150	53303
Cluster(Skoler)	1657	2515

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i lesing. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på innvandringsbakgrunn. Regresjoner er kjørt på utvalg avhengig av innvandringsbakgrunn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A11: IV-estimering med naturfagsresultater som avhengig variabel, innvandringsbakgrunn

	Innvandrere	Etnisk nordiske
Skolestørrelse	-0,007 (0,027)	0,053** (0,008)
F-verdi (førstesteg)	47,250	162,240
R^2 (førstesteg)	0,161	0,308
R^2 (andresteg)	0,194	0,229
N(Elever)	7150	53303
Cluster(Skoler)	1657	2515

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i naturfag. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på innvandringsbakgrunn. Regresjoner er kjørt på utvalg avhengig av innvandringsbakgrunn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A12: IV-estimering med leseresultater som avhengig variabel, sosioøkonomisk bakgrunn (SES)

	Lav SES	Middels SES	Høy SES
Skolestørrelse	0,059** (0,013)	0,084** (0,009)	0,101** (0,011)
F-verdi (førstesteg)	109,870	157,530	136,150
R^2 (førstesteg)	0,291	0,295	0,287
R^2 (andresteg)	0,204	0,162	0,159
N(Elever)	13302	30972	15809
Cluster(Skoler)	2378	2502	2427

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i lesing. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på antall bøker i hjemmet. Regresjoner er kjørt på utvalg avhengig av sosioøkonomisk bakgrunn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A13: IV-estimering med naturfagsresultater som avhengig variabel, sosioøkonomisk bakgrunn (SES)

	Lav SES	Middels SES	Høy SES
Skolestørrelse	0,033** (0,013)	0,046** (0,009)	0,063** (0,011)
F-verdi (førstesteg)	110,370	156,050	136,150
R^2 (førstesteg)	0,291	0,297	0,290
R^2 (andresteg)	0,229	0,175	0,151
N(Elever)	13302	30972	15809
Cluster(Skoler)	2378	2502	2427

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i naturfag. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabler på antall bøker i hjemmet. Regresjoner er kjørt på utvalg avhengig av sosioøkonomisk bakgrunn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A14: IV-estimering med leseresultater som avhengig variabel, kjønn

	Gutter	Jenter
Skolestørrelse	0,099** (0,010)	0,067** (0,009)
F-verdi (førstesteg)	154,340	157,440
R^2 (førstesteg)	0,293	0,294
R^2 (andresteg)	0,241	0,291
N(Elever)	30775	30768
Cluster(Skoler)	2499	2497

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i lesing. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabel på kjønn. Regresjoner er kjørt på utvalg avhengig av kjønn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

Tabell A15: IV-estimering med naturfagsresultater som avhengig variabel, kjønn

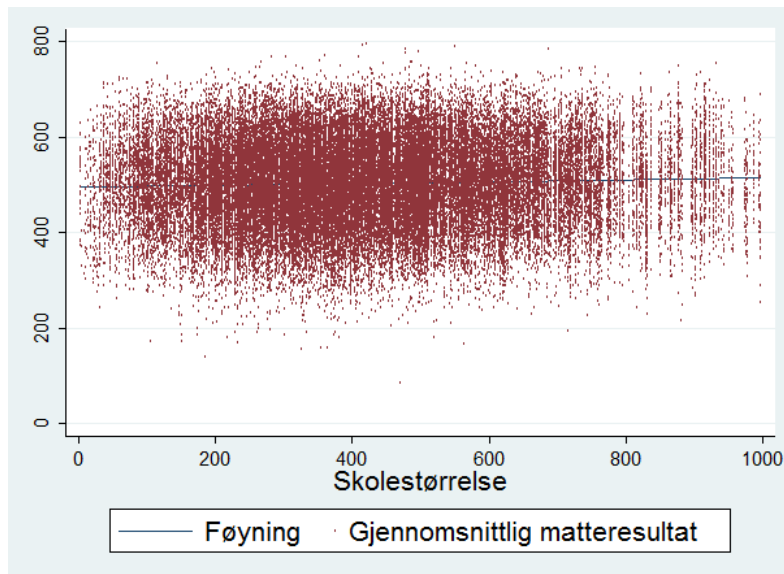
	Gutter	Jenter
Skolestørrelse	0,057** (0,001)	0,038** (0,009)
F-verdi (førstesteg)	153,070	156,910
R^2 (førstesteg)	0,295	0,295
R^2 (andresteg)	0,287	0,339
N(Elever)	30775	30768
Cluster(Skoler)	2499	2497

Signifikansnivå : † : 10% * : 5% ** : 1%

Avhengig variabel er testresultater i naturfag. Estimert med instrumentvariabel-metoden. Alle kontrollvariabler er inkludert, bortsett fra indikatorvariabel på kjønn. Regresjoner er kjørt på utvalg avhengig av kjønn som indikert i kolonneoverskrifter. Standardfeil i paranteser er korrigert for å ta høyde for clustering innen skoler.

A.3 Figurer

Figur 7: Spredningsplott med skolestørrelse og matematikkresultater hvor skolestørrelse over 1000 er fjernet



Figur 8: Sammenheng mellom skolestørrelse og matematikkresultater med kvadratisk ledd for skolestørrelse

