

Ingeborg Ottøy

ELEVPRESTASJONER OG INNVANDRERSTATUS

Mai 2020

NTNU

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

Ingeborg Ottøy

ELEVPRESTASJONER OG INNVANDRERSTATUS

Bacheloroppgave
Mai 2020

NTNU

Norges teknisk-naturvitenskapelige universitet.

Fakultet for økonomi

Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Sammendrag

Både fra et samfunnsøkonomisk perspektiv og fra et integreringsperspektiv er spørsmålet om hvordan innvandrerne klarer seg i skolesystemet viktig og aktuelt. Hensikten med denne oppgaven har vært å undersøke om det finnes en sammenheng mellom elevprestasjoner og innvandrerstatus, og hvorvidt denne sammenhengen er statistisk signifikant. Det er blitt benyttet datamateriale fra Sverige og Norge, og en sammenligning av disse to landene står også sentralt i analysen. Resultatene av den empiriske analysen tilsier at det finnes en signifikant negativ effekt av det å være innvandrer på elevprestasjonene. Effekten av å ha innvandrerforeldre er negativ og signifikant i de fleste estimerte modellene for Sverige, men man observerer ikke det samme resultatet for Norge.

Innholdsfortegnelse

Sammendrag	1
Innledning	4
<i>Motivering</i>	<i>4</i>
<i>Problemstilling.....</i>	<i>4</i>
Teoretisk rammeverk og tidligere litteratur.....	5
<i>Innledning.....</i>	<i>5</i>
<i>Teoretisk rammeverk.....</i>	<i>5</i>
<i>Tidligere funn.....</i>	<i>5</i>
<i>Oppsummering</i>	<i>7</i>
Empirisk strategi	7
<i>Innledning.....</i>	<i>7</i>
<i>Valg av funksjonsform.....</i>	<i>7</i>
<i>Estimeringsmetode.....</i>	<i>8</i>
<i>Oppsummering</i>	<i>9</i>
Datamaterialet	9
<i>Innledning.....</i>	<i>9</i>
<i>Om datamaterialet.....</i>	<i>10</i>
<i>Deskriptiv statistikk for avhengig variabel.....</i>	<i>11</i>
<i>Deskriptiv statistikk for interessevariabler.....</i>	<i>11</i>
<i>Deskriptiv statistikk for kontrollvariabler.....</i>	<i>14</i>
<i>Fordeler og begrensninger ved datamaterialet</i>	<i>15</i>
<i>Oppsummering</i>	<i>15</i>
Regresjonsanalyse og empiriske resultater.....	15
<i>Innledning.....</i>	<i>15</i>
<i>Oversikt over modellvarianter</i>	<i>16</i>
<i>Resultater: grunnmodeller</i>	<i>17</i>
<i>Resultater: ikke-lineære utvidelser</i>	<i>21</i>
Oppsummering og konklusjoner	23
<i>Hovedfunn</i>	<i>23</i>
<i>Begrensninger og mulig utvidelser.....</i>	<i>23</i>
Referanser	24
Appendiks	25

Tabeller og figurer

1: Avhengig variabel.....	10
2: Interessevariabler	10
3: Deskriptiv statistikk for avhengig variabel (Sverige)	11
4: Deskriptiv statistikk for avhengig variabel (Norge)	11
5: Deskriptiv statistikk for interessevariabler (Sverige)	11
6: Testresultater 1 (Sverige).....	12
7: Testresultater 2 (Sverige).....	12
8: Deskriptiv statistikk for interessevariabler (Norge)	12
9: Testresultater 1 (Norge).....	12
10: Testresultater 2 (Norge).....	13
11: Histogram for testresultater (Norge og Sverige)	13
12: Deskriptiv statistikk for kontrollvariabler 1.....	14
13: Deskriptiv statistikk for kontrollvariabler 2.....	14
14: Deskriptiv statistikk for kontrollvariabler 3.....	14
15: Deskriptiv statistikk for kontrollvariabler 4.....	15
16: Grunmodeller Norge.....	18
17: Grunmodeller Sverige	18
18: Hypotesetester grunnmodeller Sverige.....	20
19: Hypotesetester grunnmodeller Norge	20
20: Ikke-lineære utvidelser av modellen.....	21
21: Hypotesetester ikke-lineære utvidelser	22

Innledning

Motivering

Om lag 18% av Norges befolkning og 25% av Sveriges befolkning er enten selv innvandrere eller barn av innvandrere. Både fra et samfunnsøkonomisk perspektiv og et integreringsperspektiv er spørsmålet om hvordan innvandrerne klarer seg i skolesystemet viktig og aktuelt. I både Norge og Sverige står utdanningssystemet sentralt i integreringsprosessen, samtidig som utdanning også er svært viktig for å kunne oppnå høy sosial mobilitet i samfunnet. Studier av sammenhengen mellom elevprestasjoner og innvandrerstatus er særlig interessante fordi de kan fortelle oss noe om hvor vellykket integreringspolitikken som føres er. Det er samtidig viktig å huske på at innvandringens sammensetning og omfang har betydning for hvordan innvandrerne lykkes i det nye samfunnet de blir en del av. (SSB, 2020) (SCB, 2019)

Problemstilling

Tema for oppgaven er analyse av prestasjonsforskjeller mellom innvandrerelever og øvrige elever. Hovedproblemstillingen som undersøkes er:

- *I hvilken grad påvirker innvandrerstatus elevprestasjonene?*

Målet for den empiriske undersøkelsen er å analysere prestasjonsforskjeller *både* i Norge og Sverige og deretter foreta en sammenligning av disse funnene. Hovedtanken bak dette er at to land som er politisk og samfunnsmessig veldig like, slik som i tilfellet med de skandinaviske landene, vil gi et godt grunnlag for sammenligning av dette spesifikke fenomenet. Sverige er dog kjent for å føre den mest liberale innvandrings- og integreringspolitikken blant de skandinaviske landene, og de har som nevnt tidligere også en høyere innvandringsandel enn Norge.

Oppgaven vil også undersøke følgende tilleggsproblemstillinger:

- *Er sammenhengen mellom innvandrerstatus og elevprestasjoner betydelig større i et av de to landene?*
- *I hvilken grad avhenger betydningen av innvandrerstatus av kjønn?*

Teoretisk rammeverk og tidligere litteratur

Innledning

I dette kapitlet blir det gjort rede for det teoretiske rammeverket som den empiriske analysen baserer seg på. Videre blir det presentert noen sentrale funn om sammenhengen mellom innvandrersstatus og elevprestasjoner.

Teoretisk rammeverk

Mange av analysene som gjennomføres innenfor fagfeltet utdanningsøkonomi tar utgangspunkt i en enkel produktfunksjon. Utgangspunktet for denne oppgaven blir følgelig også at elevprestasjoner kan forklares og fremstilles ved hjelp av en slik produktfunksjon for utdanning på følgende form:

$$T = f(F, P, S)$$

F – familie/elevkarakteristika

P – medelevkarakteristika (*peer group effects*)

S – skolefaktorer

Funksjonen viser sammenhengen mellom elevenes prestasjoner, som oftest målt ved testresultater, og ulike påvirkningsfaktorer. Hovedfokuset er på de 3 gruppene forklaringsfaktorer som inngår ovenfor. Målsetningen ved å ta i bruk en skoleproduktfunksjon er vanligvis å kunne analysere og forklare forskjeller i elevprestasjoner både mellom elever, skoler og land. Sluttproduktet av utdanningsprosessen som en elev gjennomgår avhenger både av innsatsfaktorer som er direkte kontrollert av politiske bestemmelser, slik som lærerkarakteristika eller hvordan pensum er utformet, og av innsatsfaktorer som ikke er like direkte knyttet til utdanningspolitikken, slik som sosioøkonomisk bakgrunn eller grad av lærevansker. (Hanushek, 2020)

Tidligere funn

PISA (Programme for international student assessment) er en verdensomfattende undersøkelse gjennomført av OECD (organisation for economic cooperation and development) som kartlegger 15-åringers ferdigheter i lesing, naturfag og matematikk. Undersøkelsen fra 2015 ble gjennomført i 72 ulike land, og resultatene viser at innvandrere i

gjennomsnitt har dobbelt så høy sannsynlighet som øvrige elever for å prestere lavere enn basislinjenivå. I land med en høy andel innvandrere finnes det generelt ingen signifikant negativ sammenheng mellom elevprestasjoner og en høy andel innvandrere på skolen. (PISA, 2018)

Statistisk Sentralbyrå redegjør i artikkelen «Hvordan går det med innvandrere og deres barn i skolen?» for resultatene på Nasjonale Prøver i regning, lesing og engelsk fra 2016, og viser til at både innvandrere og elever med innvandrerforeldre generelt gjør det dårligere enn øvrige elever på de fleste prøvene. Unntaket var prøvene i engelsk for elever på 5.trinn. 5% av innvandrerene oppnådde det høyeste mestringsnivået i regning på 8.trinn, mens 8% av elevene med innvandrerforeldre og 11% av de øvrige elevene gjorde det samme.

SSB redegjør i samme artikkel for gjennomsnittlig antall grunnskolepoeng for de ulike gruppene. Grunnskolepoengene baserer seg på alle karakterene eleven har på vitnemålet sitt når grunnskolen er fullført. Både i 2009 og 2016 er det innvandrerene som har lavest gjennomsnittlig poengsum. Samtidig er det denne gruppen som har hatt den største økningen i gjennomsnittlig antall grunnskolepoeng fra 2009 til 2016, med en økning på 2,5 poeng mot en økning på 1,7 poeng for elevmassen samlet sett. Den største forskjellen i standpunktkarakter mellom innvandrerelver og øvrige elever for 2016 var i norsk hovedmål. Her hadde førstnevnte et gjennomsnitt på 3,2 mens snittet for øvrige elever var på 3,9. (SSB, 2017)

Grönqvist og Niknami gjør i "The school achievements of refugee children: lessons from Sweden" en analyse av skoleprestasjoner for Sveriges flyktninger. Dette innebærer at de har vurdert denne gruppen innvandrere separat. Flyktninger har utgjort en stor del av innvandringen til Sverige helt siden midten av 1980-tallet, noe som er i kontrast til mange av de andre OECD-landene. Studien baserer seg på avsluttende karakterer fra 9.klasse for perioden 1998-2014. Undersøkelsene viser at flyktningene i gjennomsnitt presterer vesentlig dårligere enn øvrige elever. Kontrollert for familiens sosioøkonomiske bakgrunn og nabolag er effekten derimot betydelig mindre. (Grönqvist og Niknami, 2017)

Oppsummering

Kapitlet startet med å presentere og forklare hovedtrekkene knyttet til teorien om skoleproduktfunksjoner. Videre har det blitt redegjort for noen sentrale funn som gir oss en indikasjon på hvilke resultater som er å forvente fra den empiriske analysen i denne oppgaven.

Empirisk strategi

Innledning

I dette kapitlet presenteres og forklares valg av funksjonsform for den empiriske analysen. Deretter redegjøres det for estimeringsmetoden og de forutsetninger som gjøres i tilknytning til dette.

Valg av funksjonsform

Basert på den teoretiske diskusjonen av skoleproduktfunksjoner i forrige kapittel så vil estimater for effekten av innvandrerstatus på elevprestasjoner bli mest mulig troverdig dersom man forsøker å kontrollere også for andre medvirkende forhold. Dette kan gjøres ved å benytte en multipl regressjonsmodell hvor den avhengige variabelen (leseprøveresultater) avhenger av en serie ulike forklaringsvariabler. Datamaterialet fra PIRLS 2001 som benyttes i denne oppgaven inneholder en rekke variabler innen de kategoriene som skoleproduktfunksjonen omfatter. Kontrollvariablene vil bli presentert mer i detalj under neste kapittel.

Grunnmodellen som danner utgangspunktet for den empiriske analysen vil være en lineær approksimasjon på følgende form:

$$(1) \textit{read} = \beta_0 + \beta_1 * \textit{not_born} + \beta_2 * \textit{par_not_born} + \beta_3 * X + \varepsilon$$

Her er variabelen «read» vårt mål på elevprestasjonene, «not_born» og «par_not_born» er dummyer for innvandrerstatus (hhv. elev født i utlandet og foreldre som er født i utlandet)

mens X angir et sett av kontrollvariabler. Epsilon utgjør det stokastiske restleddet og representerer effekten på «read» av andre variabler som ikke inngår i modellen direkte.

Estimeringsmetode

Estimeringsmetoden som benyttes er OLS, også kalt minste kvadraters metode.

Utgangspunktet for empiriske analyser er at man har et utvalg av data og ut ifra dette ønsker å *estimere* den ukjente populasjonsmodellen.

Gitt at man har en populasjonsmodell på formen:

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

Så brukes datamaterialet til å estimere en utvalgsregresjonslinje på formen:

$$\hat{Y}_i = a + bx_i$$

Hvor \hat{Y} er den predikerte verdien på den avhengige variabelen gitt x_i og anslagene a og b for alfa og beta. Den faktiske verdien på Y vil være den predikerte verdien pluss residualen:

$$Y_i = \hat{Y}_i + e_i$$

$$e_i = \hat{Y}_i - Y_i$$

Residualen e_i er det empiriske motstykket til det uobserverbare restleddet ε_i . Målet for den empiriske analysen er å estimere regresjonslinjen slik at den passer best mulig til observasjonene. Metoden OLS baserer seg på å estimere a og b slik at summen av de kvadrerte residualene minimeres:

$$\min. \sum_{i=1}^n (Y_i - a - bX_i)^2$$

Vi gjør normalt følgende forutsetninger for OLS:

- OLS-estimatorene er lineære
- Forventningen til restleddet er 0 og ingen variabler er korrelert med restleddet:
 $E(\varepsilon_i) = 0$ og $cov(\varepsilon_i, x_i) = 0$
- Variansen til restleddet er konstant: $Var(\varepsilon_i) = \sigma^2$

- De ulike observasjonene av restleddet er ikke korrelert med hverandre:
$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$
- Restleddet (epsilon) er normalfordelt. Denne forutsetningen gjøres for å kunne gjennomføre hypotesetester.

Under disse forutsetningene er OLS-estimatorene forventningsrette og effisiente. Dette innebærer at estimatoren ikke avviker systematisk fra den sanne parameterverdien, samt at det ikke finnes estimatører med lavere varians (vi oppnår høyest mulig presisjon).

Hovedinteressen for den empiriske analysen er estimering av parameterne beta1 og beta2 i likning (1) og undersøkelser av disse variablenes signifikans for elevprestasjonene.

Utgangspunktet er en grunnmodell slik den er beskrevet ovenfor, og modellen utvides ved at ulike sett med kontrollvariabler inkluderes. Dette gjør det mulig å undersøke hvordan effekten av innvandrerststatus endres, og hvorvidt den er robust, når vi utvider modellen. Tilleggsproblemstillingene som er presentert i innledningen undersøkes ved å inkludere interaksjonsledd i modellen.

Oppsummering

Det har i dette kapittelet vært redegjort for valget av en lineær approksimasjon som funksjonsform. Videre er estimeringsmetoden OLS blitt presentert og forklart. Til slutt ble fremgangsmåten for oppgavens empiriske analyse presentert i korte trekk.

Datamaterialet

Innledning

I dette kapittelet redegjøres det kort for datamaterialet og variabler som benyttes i de økonometriske modellene. Deretter presenteres sentral deskriptiv statistikk for de aktuelle variablene. Til slutt diskuteres fordeler og begrensninger ved materialet.

Om datamaterialet

Datamaterialet som benyttes i denne oppgaven er et tilrettelagt materiale hentet fra den fullstendige undersøkelsen Progress in International Reading Literacy Study (PIRLS) fra 2001. Dette er en internasjonal undersøkelse gjennomført i 35 ulike land med formål om å teste leseferdigheter til elever i fjerde klasse. Det var International association for the evaluation of educational achievement (IEA) som sto for gjennomførelsen og organiseringen av undersøkelsen. Undersøkelsen besto av en skriftlig test for leseforståelse som elevene gjennomførte, samt at elever, lærere, skolens ledelse og foreldre svarte på et spørreskjema. Her ble det innhentet informasjon om faktorer knyttet til hjem og skole som er av betydning for utviklingen av leseferdigheter.

Definisjon av variabler som benyttes i analysen:

Avhengig variabel

1: Avhengig variabel

Navn	Innhold	Type variabel
Read	Testresultat leseprøve	Kontinuerlig

Interessevariabler

2: Interessevariabler

Navn	Innhold	Verdi	Type variabel
Not_born	Elev født i landet	0	Dummy
	Elev født i utlandet	1	
Par_not_born	Foreldre født i landet	0	Dummy
	Foreldre født i utlandet	1	

Kontrollvariabler

Se appendiks for oversikt over benyttede sett med kontrollvariabler. Kategorivariabler er omgjort til passende kvalitative variabler (dummyer) som benyttes i de økonometriske modellene. Mange steder er flere av de opprinnelige kategoriene slått sammen slik at man unngår et svært høyt antall variabler. Legg merke til at én av kategoriene brukes som referansekategori i regresjonen slik at problemer med perfekt multikollinearitet unngås. Multikollinearitet er graden av lineær sammenheng mellom forklaringsvariablene i en

multippel regresjonsmodell, og i dette tilfellet ville det oppstått en eksakt lineær sammenheng mellom dummyvariablene.

Deskriptiv statistikk for avhengig variabel

Sverige

3: Deskriptiv statistikk for avhengig variabel (Sverige)

	READ
Gjennomsnitt	564,71
Standardavvik	61,31
Maksimumsverdi	737,33
Minimumsverdi	318,68
Antall observasjoner	7 199

Norge

4: Deskriptiv statistikk for avhengig variabel (Norge)

	READ
Gjennomsnitt	498,26
Standardavvik	78,37
Maksimumsverdi	695,87
Minimumsverdi	228,06
Antall observasjoner	3 459

Vi ser fra dette at antall observasjoner er betydelig større for Sverige, noe som også er logisk da Sverige har flere innbyggere enn Norge. Videre er det verdt å merke seg at gjennomsnittlig testresultat også er betydelig høyere for de svenske elevene.

Deskriptiv statistikk for interessevariabler

Sverige

5: Deskriptiv statistikk for interessevariabler (Sverige)

	NOT_BORN	PAR_NOT_BORN
Gjennomsnitt	0,121	0,123
Standardavvik	0,326	0,328
Maksimumsverdi	1	1
Minimumsverdi	0	0
Antall observasjoner	7 028	6 988

Testresultater for henholdsvis innvandrere og øvrige elever:

6: Testresultater 1 (Sverige)

	READ IF NOT_BORN=1	READ IF NOT_BORN=0
Gjennomsnitt	524,39	570,22
Standardavvik	66,45	58,69
Maksimumsverdi	709,99	737,33
Minimumsverdi	318,68	340,55
Antall observasjoner	851	6 177

Testresultater for henholdsvis elever med innvandrerforeldre og øvrige elever:

7: Testresultater 2 (Sverige)

	READ IF PAR_NOT_BORN=1	READ IF PAR_NOT_BORN=0 & NOT_BORN=0
Gjennomsnitt	529,13	572,91
Standardavvik	65,25	57,77
Maksimumsverdi	709,99	737,33
Minimumsverdi	318,68	340,55
Antall observasjoner	858	5 622

Norge

8: Deskriptiv statistikk for interessevariabler (Norge)

	NOT_BORN	PAR_NOT_BORN
Gjennomsnitt	0,091	0,058
Standardavvik	0,288	0,240
Maksimumsverdi	1	1
Minimumsverdi	0	0
Antall observasjoner	3 355	3 374

Testresultater for henholdsvis innvandrere og øvrige elever:

9: Testresultater 1 (Norge)

	READ IF NOT_BORN=1	READ IF NOT_BORN=0
Gjennomsnitt	463,79	502,76
Standardavvik	83,41	76,69
Maksimumsverdi	673,03	695,87
Minimumsverdi	234,89	228,06
Antall observasjoner	305	3 050

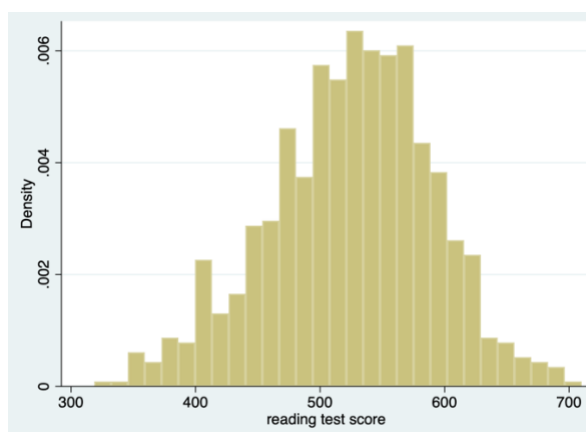
Testresultater for henholdsvis elever med innvandrerforeldre og øvrige elever:

10: Testresultater 2 (Norge)

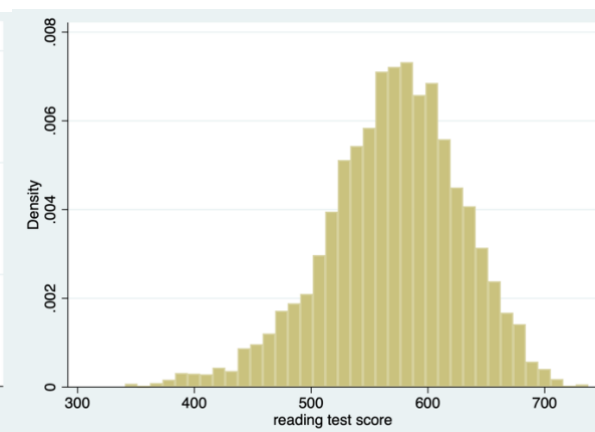
	READ IF PAR_NOT_BORN=1	READ IF PAR_NOT_BORN=0 & NOT_BORN=0
Gjennomsnitt	444,34	504,95
Standardavvik	87,84	75,58
Maksimumsverdi	636,77	695,87
Minimumsverdi	234,89	228,06
Antall observasjoner	197	2 919

Histogrammer som viser tetthet:

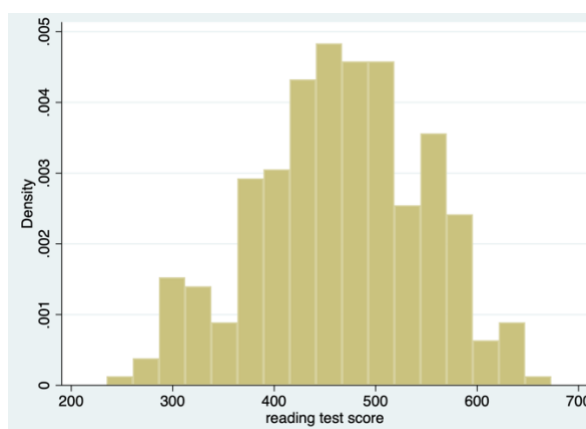
11: Histogram for testresultater (Norge og Sverige)



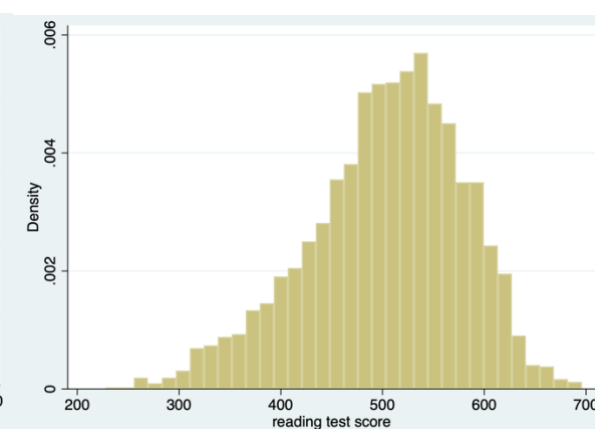
Innvandrererelever i Sverige



Elever født i Sverige



Innvandrererelever i Norge



Elever født i Norge

Gjennomsnittsverdiene til dummyene for innvandrerstatus og innvandrerforeldre viser at andelen med innvandrerbakgrunn er høyere i Sverige enn i Norge, noe som også er i tråd med forventningene. Gjennomsnittlig testresultater er lavere for innvandrere og elever med innvandrerforeldre sammenlignet med øvrige elever, både for Norge og Sverige.

Deskriptiv statistikk for kontrollvariabler

Her presenteres deskriptiv statistikk for de ulike gruppene av kontrollvariabler samlet for Norge og Sverige. Den første gruppen kontrollvariabler som inkluderes i modellen er en samling av individ- og familiekarakteristika. De neste to gruppene er lærerkarakteristika og klassekarakteristika, som begge inngår i hovedgruppen skolefaktorer. Til slutt inkluderes en gruppe kontrollvariabler, kalt skolekarakteristika, som omfatter både variabler for skolefaktorer og medelevkarakteristika.

Individ- og familiekarakteristika

12: Deskriptiv statistikk for kontrollvariabler 1

	Gjennomsnitt	Std.avvik	Max.	Min.	Ant. obs.
Girl	0,49	0,5	1	0	10 465
Dtestlang_always	0,88	0,32	1	0	10 417
Dtestlang_sometimes	0,098	0,3	1	0	10 417
Dbooks100plus	0,71	0,46	1	0	10 658
Dincome40plus	0,69	0,46	1	0	10 658
Dedu_higher	0,59	0,49	1	0	9 601
Demp_full	0,95	0,22	1	0	9 147

Lærerkarakteristika

13: Deskriptiv statistikk for kontrollvariabler 2

	Gjennomsnitt	Std.avvik	Max.	Min.	Ant. obs.
Dteacher40plus	0,64	0,48	1	0	10 658
Teacher_exp	15,77	11,31	42	1	10 078
Teacher_fem	0,83	0,38	1	0	10 306
Teacher_cert	0,95	0,23	1	0	10 250

Klassekarakteristika

14: Deskriptiv statistikk for kontrollvariabler 3

	Gjennomsnitt	Std.avvik	Max.	Min.	Ant. obs.
Clsize	24,71	8,29	57	4	10 255

Skolekarakteristika

15: Deskriptiv statistikk for kontrollvariabler 4

	Gjennomsnitt	Std.avvik	Max.	Min.	Ant. obs.
Schoolsize4	55,82	32,65	192	3	10 142
Dlocation100plus	0,35	0,48	1	0	10 658
Ddisadv25plus	0,23	0,42	1	0	10 658
Dabroad25plus	0,11	0,31	1	0	10 658

Fordeler og begrensninger ved datamaterialet

Datamaterialet inneholder et høyt antall observasjoner, noe som gjør det godt egnet til å drive regresjonsanalyse. Materialet inneholder også omfattende informasjon om ulike forhold som kan tenkes å påvirke testresultatene, slik at man har gode muligheter til å kontrollere for slike medvirkende forhold i de økonometriske modellene. Testresultatene som danner grunnlaget for undersøkelsen er knyttet til leseferdigheter, og omfatter ikke andre typer sentrale ferdigheter slik som f.eks regning. Man kan stille spørsmål ved om resultatene av analysen da kan generaliseres til elevprestasjoner generelt, eller om dette vil være problematisk.

Oppsummering

I dette kapitlet har datamaterialet som benyttes i analysen blitt presentert. Deretter har det blitt presentert hovedtrekk og sentrale egenskaper ved de aktuelle variablene. Til slutt ble fordeler og begrensninger ved datamaterialet diskutert.

Regresjonsanalyse og empiriske resultater

Innledning

I dette kapitlet gis det først en oversikt over de ulike modellvariantene som har blitt estimert. Deretter følger en presentasjon og gjennomgang av resultatene og de hypotesetestene som er gjennomført. Resultater knyttet til grunnmodellene presenteres først, deretter resultater knyttet til de ikke-lineære utvidelsene.

Oversikt over modellvarianter

$$(1) \textit{read} = \beta_0 + \beta_1 * \textit{not_born} + \beta_2 * \textit{par_not_born} + \beta_3 * X + \varepsilon$$

Med utgangspunkt i denne lineære modellen estimeres følgende 5 varianter for Norge og Sverige separat:

- (1) En enkel regresjon med interessevariablene som eneste forklaringsvariabler
- (2) Grunnmodell hvor X består av et sett med kontrollvariabler for individ- og familiekarakteristika
- (3) Grunnmodell hvor X nå utvides med kontrollvariabler for lærerkarakteristika
- (4) Grunnmodell hvor X nå utvides med kontrollvariabler for klassekarakteristika
- (5) Grunnmodell hvor X nå utvides med kontrollvariabler for skolekarakteristika

Første tilleggsproblemstilling undersøkes ved å estimere en ikke-lineær utvidelse av modellen på følgende form:

$$(2) \textit{read} = \beta_0 + \beta_1 * \textit{not_born} + \beta_2 * \textit{par_not_born} + \beta_3 * X + \beta_4 * \textit{DSverige} + \beta_5 * \textit{Sv_not_born} + \beta_6 * \textit{Sv_par_not_born}$$

Modellen estimeres basert på hele datasettet (Norge og Sverige), og inkluderer derfor både en dummy for Sverige og de to følgende interaksjonsleddene:

- Interaksjonsledd mellom dummy for Sverige og dummy for innvandrerstatus (*Sv_not_born*)
- Interaksjonsledd mellom dummy for Sverige og dummy for innvandrerforeldre (*Sv_par_not_born*)

Andre tilleggsproblemstilling undersøkes ved å estimere en ikke-lineær utvidelse av modellen på følgende form:

$$(3) \textit{read} = \beta_0 + \beta_1 * \textit{not_born} + \beta_2 * \textit{par_not_born} + \beta_3 * X + \beta_4 * \textit{girl_not_born} + \beta_5 * \textit{girl_par_not_born}$$

Modellen estimeres for Norge og Sverige separat, og inkluderer følgende to interaksjonsledd:

- Interaksjonsledd mellom dummy for jente og dummy for innvandrerstatus (girl_not_born)
- Interaksjonsledd mellom dummy for jente og dummy for innvandrerforeldre (girl_par_not_born)

Resultater: grunnmodeller

Fullstendige tabeller for regresjonene finnes i appendiks. Her presenteres oppsummerende tabeller som viser hovedresultatene for den empiriske analysen.

Kommentarer til tabellene:

Det øverste av de to tallene i tabellen er den estimerte verdien på koeffisientene (β_k). Tallet som står i parentes angir standardavviket til disse estimatorene og er gitt som:

$$s_{bk} = \frac{s^2}{\sum(X_{ki} - \bar{x}_k)(1 - r_{kj}^2)}$$

k: nr. på variabel

i: observasjon nr. i av variabelen *k*

s^2 : forventningsrett estimator for populasjonsvariansen

\bar{x}_k : gjennomsnittlig verdi for variabel *k*

r_{kj}^2 : korrelasjon mellom variabel *k* og øvrige variabler

R^2 angir modellens forklaringskraft. Denne sier noe om hvor stor andel av den totale variasjonen i den avhengige variabelen som forklares av modellen.

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

SSE: forklart varians

SST: total varians

Oppsummerende tabell: Grunnmodeller for Norge

16: Grunnmodeller Norge

	(1) Read	(2) Read	(3) Read	(4) Read	(5) Read
Not_born (X1)	-25.39 (4.867)	-25.70 (5.136)	-26.97 (5.207)	-27.07 (5.209)	-25.79 (5.384)
Par_not_born (X2)	-49.25 (5.911)	-3.637 (7.232)	-1.737 (7.405)	-1.867 (7.406)	-0.449 (7.769)
Individ- og familiekarakteristika	Nei	Ja	Ja	Ja	Ja
Lærerkarakteristika	Nei	Nei	Ja	Ja	Ja
Klassekarakteristika	Nei	Nei	Nei	Ja	Ja
Skolekarakteristika	Nei	Nei	Nei	Nei	Ja
Observasjoner (N)	3328	2805	2706	2706	2534
R ²	0.039	0.156	0.158	0.158	0.161

Standardavvik oppgitt i parentes

Oppsummerende tabell: Grunnmodeller for Sverige

17: Grunnmodeller Sverige

	(1) Read	(2) Read	(3) Read	(4) Read	(5) Read
Not_born (X1)	-35.97 (2.305)	-31.84 (2.459)	-30.31 (2.568)	-30.19 (2.584)	-31.03 (2.612)
Par_not_born (X2)	-28.94 (2.282)	-7.637 (3.100)	-7.442 (3.233)	-7.143 (3.250)	-4.557 (3.343)
Individ- og familiekarakteristika	Nei	Ja	Ja	Ja	Ja
Lærerkarakteristika	Nei	Nei	Ja	Ja	Ja
Klassekarakteristika	Nei	Nei	Nei	Ja	Ja
Skolekarakteristika	Nei	Nei	Nei	Nei	Ja
Observasjoner (N)	6965	5906	5488	5454	5205
R ²	0.080	0.154	0.157	0.155	0.174

Standardavvik oppgitt i parentes

Kommentarer til regresjonene

- Vi ser at alle 5 modellene estimerer negativ effekt på testresultatene både for innvandrerelever og elever med innvandrerforeldre. I modell (5) vil en elev som er innvandrer i Norge kunne forvente et resultat som er 25,79 poeng lavere enn for øvrige elever. Tilsvarende vil en elev som er innvandrer i Sverige kunne forvente et resultat som er 31,03 poeng lavere enn øvrige elever. Dette er en reduksjon i poengsum på hhv. omtrent 7% og 6% av modellenes konstantledd. Koeffisientene for de andre modellene kan tolkes på tilsvarende måte.
- Effekten av å være innvandrer selv er relativt stabil for Norge, uavhengig av hvilke sett med kontrollvariabler som inkluderes.
- Effekten av å være innvandrer selv reduseres fra -35,97 til -31,84 for Sverige når kontrollvariabler for individ- og familiekarakteristika inkluderes. Videre forblir effekten relativt stabil.
- For begge land reduseres den negative effekten av å ha innvandrerforeldre kraftig når den enkle modellen (1) utvides til å inkludere også individ- og familiekarakteristika (2). Dette tyder på at den første modellen har et betydelig utelatt variabel-problem. Et slik problem oppstår dersom de utelatte variablene korrelerer med de inkluderte variablene. Koeffisienten til den inkluderte variabelen vil da i realiteten vise en brutto-effekt som omfatter både den direkte effekten av inkludert variabel på avhengig variabel og den indirekte effekten av de utelatte variablene via den inkluderte variabelen. En slik feilspesifikasjon av modellen gir derfor skjeve OLS-estimatorer. Det er i tråd med forventningene at det finnes en samvariasjon mellom det å ha innvandrerforeldre og det å ha en ufordelaktig sosioøkonomisk bakgrunn.

Hypotesetester

Det har blitt tatt i bruk t-test for å teste hypoteser om hvorvidt effekten av innvandrerstatus er statistisk signifikant for de ulike modellene og for hvert land. Gitt de forutsetningene som er lagt til grunn for OLS-regresjonen så vil estimatorene være normalfordelte:

$$b_k \sim N(\mu_{\beta_k}, \sigma_{\beta_k}^2) \quad k = 1, 2, \dots, n$$

Hypotesetestene settes opp på følgende form:

$$H_0: \beta_k = 0$$

$$H_A: \beta_k \neq 0$$

Dette gir oss følgende t-fordelte testobservator med $(n-k)$ frihetsgrader, hvor populasjonsstandardavviket erstattes med sin forventningsrette estimator (s):

$$TS = \frac{b_k - 0}{s_k} \sim t(n - k)$$

H_0 forkastes dersom den absolutte verdien på testobservatoren er større enn den kritiske verdien for t-fordelingen, gitt antall frihetsgrader.

Sverige

18: Hypotesetester grunnmodeller Sverige

Modell	Testobservator Not_born	Testobservator Par_not_born	Konklusjon
(1)	-15.61	-12.68	Forkast // forkast
(2)	-12.95	-2.46	Forkast // forkastes på 5%
(3)	-11.81	-2.30	Forkast // forkastes på 5%
(4)	-11.68	-2.20	Forkast // forkastes på 5%
(5)	-11.88	-1.36	Forkast // ikke forkast

Norge

19: Hypotesetester grunnmodeller Norge

Modell	Testobservator Not_born	Testobservator Par_not_born	Konklusjon
(1)	-5.22	-8.33	Forkast // forkast
(2)	-5.00	-0.50	Forkast // ikke forkast
(3)	-5.18	-0.23	Forkast // ikke forkast
(4)	-5.20	-0.25	Forkast // ikke forkast
(5)	-4.79	-0.06	Forkast // ikke forkast

Undersøkelsene viser at den negative effekten på elevprestasjonene av å være innvandrere er statistisk signifikant på alle signifikansnivå, for alle modellene og for begge land. For Norge er effekten på elevprestasjonene av å ha innvandrerforeldre kun statistisk signifikant for den første regresjonen, gitt at man benytter seg av konvensjonelle signifikansnivå. Dette skyldes trolig utelatt variabelproblemet som har blitt diskutert tidligere. For Sverige er derimot denne effekten statistisk signifikant på et 5% signifikansnivå for modellene (2)-(4), men ikke signifikant for den siste grunnmodellen.

Resultater: ikke-lineære utvidelser

Modell (1) er knyttet til undersøkning av hypotesen om hvorvidt effekten av innvandrerstatus er større for et av de to landene. Modell (2) og (3) er knyttet til undersøkning av hypotesen om hvorvidt effekten av innvandrerstatus avhenger av kjønn.

Oppsummerende tabell

20: Ikke-lineære utvidelser av modellen

	(1) Read	Sverige (2) Read	Norge (3) Read
Not_born (X1)	-27.41 (4.571)	-30.13 (3.465)	-33.87 (7.507)
Par_not_born (X2)	-11.50 (6.170)	-0.438 (4.220)	0.251 (10.27)
Individ- og familiekarakteristika	Ja	Ja	Ja
Lærerkarakteristika	Ja	Ja	Ja
Klassekarakteristika	Ja	Ja	Ja
Skolekarakteristika	Ja	Ja	Ja
DSverige	63.25 (1.724)	Nei	Nei
Sv_not_born	-2.800 (5.399)	Nei	Nei
Sv_par_not_born	9.049 (6.709)	Nei	Nei
Girl_not_born	Nei	-1.972	16.54

		(5.241)	(10.68)
Girl_par_not_born	Nei	-8.675 (5.416)	-0.597 (13.94)
Observasjoner (N)	7739	5205	2534
R ²	0.310	0.174	0.162

Standardavvik oppgitt i parentes

Kommentarer til regresjonene

- Modell (1) viser, som forventet, en relativt sterk positiv effekt på testresultatene av å være svensk (63,25 poeng). Videre estimeres det en større negativ effekt av å være innvandrerelev dersom man er svensk. Effekten av å være innvandrer vil da være hhv. $(-27,41-2,8) = \underline{-30,21}$ i Sverige og $\underline{-27,41}$ i Norge. For elever med innvandrerforeldre er derimot den estimerte effekten motsatt, med $(-11,50+9,049) = \underline{-2,45}$ for Sverige og $\underline{-11,50}$ for Norge.
- Modell (2) og (3) viser at effekten av innvandrerstatus avhenger ulikt av kjønn i de 2 landene. For Sverige estimeres det at en innvandrerelev eller elev med innvandrerforeldre som også er jente vil ha dårligere testresultater enn en gutt. Modellen for Norge estimerer derimot en sterk positiv effekt av å være jente dersom man selv er innvandrer (16,54 poeng).

Hypotesetester

21: Hypotesetester ikke-lineære utvidelser

Variabel	Testobservator	Konklusjon
Sv_not_born	-0.52	Ikke forkast
Sv_par_not_born	1.35	Ikke forkast

Variabel	Testobservator Sverige (2)	Testobservator Norge (3)	Konklusjon
Girl_not_born	-0.38	1.55	Ikke forkast // ikke forkast
Girl_par_not_born	-1.60	-0.04	Ikke forkast // ikke forkast

Estimatene for ulik effekt av innvandrerstatus i Norge og Sverige er ikke statistisk signifikante på konvensjonelle signifikansnivå. Dette er også tilfellet for effekten av kjønn. Det er likevel aktuelt å peke på at den negative effekten av å være jente dersom man har innvandrerforeldre er signifikant helt ned på et 10,9% signifikansnivå for Sverige. I tillegg er den positive effekten av å være jente dersom man er innvandrerrelatert signifikant helt ned på et 12,2% signifikansnivå for Norge. Disse to resultatene er verdt å legge merke til.

Oppsummering og konklusjoner

Hovedfunn

Den empiriske analysen finner en negativ sammenheng mellom elevprestasjoner og det å være innvandrer som er signifikant for alle modellene, for begge land og på alle signifikansnivå. For Norge finner man ingen statistisk signifikant negativ effekt av å ha innvandrerforeldre på elevprestasjonene når det kontrolleres for andre medvirkende forhold. For Sverige er derimot denne sammenhengen statistisk signifikant på et 5% signifikansnivå når det kontrolleres for individ- og familiekarakteristika, lærerkarakteristika og klassekarakteristika. De ikke-lineære modellspesifikasjonene gir ingen statistisk signifikante resultater når man benytter konvensjonelle signifikansnivå.

Begrensninger og mulig utvidelser

- Analysen baserer seg kun på leseferdigheter, slik at det kan være problematisk å generalisere til elevprestasjoner generelt. Mulig relevant utvidelse er å benytte testresultater også for andre typer ferdigheter.
- Datamaterialet gir oss ikke informasjon om innvandringens sammensetting. En mulig relevant utvidelse er å benytte informasjon om hvor innvandrerne kommer fra og hva som er årsaken.
- Analysen tar kun utgangspunkt i Norge og Sverige. Det vil trolig være problematisk å generalisere disse resultatene til andre land med andre institusjonelle- og samfunnsmessige forhold.

Referanser

Grönqvist, Hans og Niknami, Susan (2017) The school achievements of refugee children: lessons from Sweden. *Nordic economic policy review*, s. 159 – 184.

<http://www.diva-portal.org/smash/get/diva2:1090694/FULLTEXT01.pdf#page=161>

Hanushek, Eric (2020) *Education production functions*. I Steve Bradley og Colin Green (red.), *Economics of Education*, 2nd Edition, London: Academic Press, 2020, s. 161-170.

<http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202020%20Education%20Production%20Functions.pdf>

IEA *Progress in international reading literacy study 2001*. Lokalisert på internet 14.04.2020 fra:

<https://www.iea.nl/studies/iea/pirls/2001>

PISA (2018) *PISA 2015 Results in focus*. Lokalisert på internett 27.04.2020 fra:

<https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>

SCB (2019) *Sveriges befolkning efter födelseland, medborgarskap och bakgrund*. Lokalisert på internett 26.04.2020 fra:

<https://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningens-sammansattning/befolkningsstatistik/pong/tabell-och-diagram/helarsstatistik--riktet/sveriges-befolkning-efter-fodelseland-region-medborgarskap-och-bakgrund-31-december-2019/>

SSB (2020) *Faktaside innvandring*. Lokalisert på internett 05.04.2020 fra:

<https://www.ssb.no/innvandring-og-innvandrere/faktaside/innvandring>

SSB (2017) *Hvordan går det med innvandrere og deres barn i skolen?* Lokalisert på internett 13.04.2020 fra:

<https://www.ssb.no/utdanning/artikler-og-publikasjoner/hvordan-gar-det-med-innvandrere-og-deres-barn-i-skolen>

Appendiks

Definisjoner av kontrollvariabler

Navn	Innhold	Type variabel	Erstatter kategorivariabel
Girl	Eleven er jente	Dummy	
Dtestlang_always	Snakker testspråket hjemme alltid	Dummy	Speak_testlang_home
Dtestlang_sometimes	Snakker testspråket hjemme noen ganger	Dummy	Speak_testlang_home
Dbooks100plus	Mer enn 100 bøker i hjemmet	Dummy	Books_home
Dincome40plus	Husholdningens inntekt er 40 000 dollar eller mer	Dummy	Income
Dedu_higher	Foreldre har høyere utdanning	Dummy	Par_edu
Demp_full	En eller begge foreldre arbeider fulltid	Dummy	Par_emp
Teacher_exp	Antall år erfaring som lærer	Kontinuerlig	
Teacher_fem	Lærer er kvinne	Dummy	
Teacher_cert	Lærer er sertifisert	Dummy	
Dteacher40plus	Lærer er 40 år eller eldre	Dummy	Teacher_age
Clsize	Antall elever i klassen	Kontinuerlig	
Schoolsize4	Antall 4.klassinger på skolen	Kontinuerlig	
Dlocation100plus	Skoledistrikt har mer enn 100 000 innbyggere	Dummy	School_location
Ddisadv25plus	Mer enn 25% av elevene kommer fra vanskeligstilte hjem	Dummy	Pct_disadv
Dabroad25plus	Mer enn 25% av elevene er født i utlandet	Dummy	Pct_abroad

Fullstendig tabell for grunnmodeller Sverige:

	(1) read	(2) read	(3) read	(4) read	(5) read
not_born	-35.97 (2.305)	-31.84 (2.459)	-30.31 (2.568)	-30.19 (2.584)	-31.03 (2.612)
par_not_born	-28.94 (2.282)	-7.637 (3.100)	-7.442 (3.233)	-7.143 (3.250)	-4.557 (3.343)
girl		20.96 (1.424)	21.19 (1.474)	21.28 (1.480)	20.04 (1.494)
Dtestlang_always		12.61 (6.549)	17.13 (6.826)	16.91 (6.990)	18.43 (6.947)
Dtestlang_someti mes		5.339 (6.559)	9.814 (6.835)	10.13 (7.007)	9.688 (6.971)
Dbooks100plus		15.98 (1.675)	15.45 (1.735)	15.50 (1.743)	11.98 (1.772)
Dincome40plus		6.731 (1.598)	6.242 (1.656)	6.059 (1.666)	5.377 (1.690)
Dedu_higher		22.41 (1.576)	22.40 (1.629)	22.31 (1.637)	20.42 (1.665)
Demp_full		3.766 (3.365)	4.203 (3.461)	4.623 (3.491)	6.273 (3.604)
teacher_exp			0.251 (0.0991)	0.237 (0.100)	0.197 (0.101)
teacher_fem			6.501 (1.873)	6.297 (1.893)	5.705 (1.883)
teacher_cert			3.960 (3.079)	3.673 (3.105)	1.820 (3.082)
Dteacher40plus			-0.558 (2.261)	-0.447 (2.285)	2.549 (2.306)
clsize				0.103 (0.0839)	0.198 (0.0877)
schoolsize4					0.0864 (0.0220)
Dlocation100plus					13.35 (1.767)
Ddisadv25plus					0.105 (1.963)

Dabroad25plus					-17.96 (2.916)
_cons	572.8 (0.777)	518.3 (7.236)	501.5 (8.180)	499.2 (8.440)	492.9 (8.556)
<i>N</i>	6965	5906	5488	5454	5205
<i>R</i> ²	0.080	0.154	0.157	0.155	0.174

Standardavvik er oppgitt i parentes

Fullstendig tabell for grunnmodeller Norge:

	(1) read	(2) read	(3) read	(4) read	(5) read
not_born	-25.39 (4.867)	-25.70 (5.136)	-26.97 (5.207)	-27.07 (5.209)	-25.79 (5.384)
par_not_born	-49.25 (5.911)	-3.637 (7.232)	-1.736 (7.405)	-1.867 (7.406)	-0.449 (7.769)
girl		21.13 (2.650)	20.41 (2.700)	20.37 (2.700)	20.25 (2.786)
Dtestlang_always		68.12 (14.14)	75.65 (14.46)	75.75 (14.46)	79.46 (15.29)
Dtestlang_someti mes		50.36 (14.36)	57.21 (14.65)	57.19 (14.65)	62.60 (15.53)
Dbooks100plus		23.84 (3.125)	23.23 (3.175)	23.23 (3.175)	22.77 (3.289)
Dincome40plus		5.653 (3.033)	5.271 (3.078)	5.056 (3.086)	4.213 (3.191)
Dedu_higher		34.18 (2.917)	34.41 (2.966)	34.16 (2.977)	34.38 (3.078)
Demp_full		9.409 (6.522)	11.14 (6.644)	11.34 (6.647)	13.23 (6.887)
teacher_exp			-0.161 (0.160)	-0.158 (0.160)	-0.195 (0.170)
teacher_fem			2.118 (4.062)	2.329 (4.068)	1.490 (4.352)
teacher_cert			-2.670 (8.012)	-2.218 (8.025)	3.913 (8.366)
Dteacher40plus			-1.701 (3.848)	-1.585 (3.850)	-1.645 (4.114)
clsize				0.269	0.198

				(0.277)	(0.322)
schoolsize4					0.0866 (0.0792)
Dlocation100plus					5.711 (2.947)
Ddisadv25plus					-1.237 (4.734)
Dabroad25plus					8.816 (7.385)
_cons	504.7 (1.406)	383.0 (15.17)	379.1 (18.04)	372.8 (19.18)	358.9 (20.39)
<i>N</i>	3328	2805	2706	2706	2534
<i>R</i> ²	0.039	0.156	0.158	0.158	0.161

Standardavvik er oppgitt i parentes

Fullstendig tabell for ikke-lineære modeller:

	(1) read	(2) read	(3) read
not_born	-27.41 (4.571)	-30.13 (3.465)	-33.87 (7.507)
par_not_born	-11.50 (6.170)	-0.438 (4.220)	0.251 (10.27)
DSverige	63.25 (1.724)		
girl	20.23 (1.362)	21.02 (1.610)	18.94 (2.931)
Dtestlang_always	34.70 (6.650)	18.38 (6.950)	80.28 (15.29)
Dtestlang_someti mes	23.19 (6.698)	9.739 (6.971)	63.09 (15.53)
Dbooks100plus	15.93 (1.614)	11.98 (1.772)	22.79 (3.289)
Dincome40plus	5.498 (1.545)	5.338 (1.690)	4.193 (3.191)
Dedu_higher	25.62 (1.510)	20.44 (1.665)	34.44 (3.078)
Demp_full	8.720 (3.315)	6.207 (3.604)	13.35 (6.897)

teacher_exp	0.0554 (0.0871)	0.195 (0.101)	-0.194 (0.170)
teacher_fem	4.392 (1.797)	5.792 (1.883)	1.477 (4.351)
teacher_cert	3.248 (3.042)	1.832 (3.082)	4.079 (8.369)
Dteacher40plus	1.381 (2.041)	2.619 (2.307)	-1.562 (4.114)
clsize	0.239 (0.0896)	0.195 (0.0877)	0.205 (0.322)
schoolsize4	0.0908 (0.0226)	0.0867 (0.0220)	0.0841 (0.0792)
Dlocation100plus	9.728 (1.518)	13.30 (1.767)	5.473 (2.951)
Ddisadv25plus	1.121 (1.905)	0.170 (1.963)	-1.030 (4.739)
Dabroad25plus	-10.03 (2.853)	-17.90 (2.917)	8.907 (7.392)
Sv_not_born	-2.800 (5.398)		
Sv_par_not_born	9.049 (6.709)		
girl_not_born		-1.972 (5.241)	16.54 (10.68)
girl_par_not_born		-8.675 (5.416)	-0.597 (13.94)
_cons	406.2 (8.238)	492.5 (8.558)	358.4 (20.39)
<i>N</i>	7739	5205	2534
<i>R</i> ²	0.309	0.174	0.162

Standardavvik er oppgitt i parentes