

Bendik Barlinn Kjelstad
Ella Johanne Devold
Jonatan Skevik

Hvordan avhenger elevprestasjoner i Norge og Brasil av sosial klasse?

Bacheloroppgave i Samfunnsøkonomi
Veileder: Sepehr Cyrusian
Mai 2024

Bendik Barlinn Kjelstad
Ella Johanne Devold
Jonatan Skevik

Hvordan avhenger elevprestasjoner i Norge og Brasil av sosial klasse?

Bacheloroppgave i Samfunnsøkonomi
Veileder: Sepehr Cyrusian
Mai 2024

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

1 Forord

Følgende bacheloroppgave er skrevet som en konkluderende oppgave for vårt bachelorstudium innen samfunnsøkonomi ved Fakultet for Økonomi ved Norges teknisk-naturvitenskapelige universitet.

Vi ønsker å uttrykke stor takknemlighet overfor vår veileder, Sepehr Cyrusian (Stipendiat, Institutt for Samfunnsøkonomi), for hans veiledning og støtte gjennom arbeidet med oppgaven.

Bendik B. Gjølstad

Eivind Devold

Sepehr

2 Sammendrag

Flere tidligere studier peker på hvordan elevprestasjoner avhenger av elevens sosioøkonomiske status. Denne oppgaven ser på hvordan leseferdigheter hos 4. klasse-elever påvirkes av foreldrenes inntekt og utdanningsnivå, samt tilgangen på bøker og barnebøker i hjemmet. Vi har valgt å undersøke dette for Norge og Brasil, for å undersøke nærmere om disse effektene vil ha større betydning i land med store sosiale forskjeller. Gjennom analysen observerer vi at leseferdighetene varierer avhengig av sosial klasse, og at denne effekten er større i land med store sosiale forskjeller. Vi kontrollerer også for kjønn, og observerer at gutter opplever denne effekten sterkere enn jenter. I denne oppgaven gjør vi flere typer tester for å underbygge konklusjonene nevnt over. Avslutningsvis presenterer vi analysens begrensninger, og peker på potensialet for forbedringer og videre arbeid.

Innhold

1 Forord	i
2 Sammendrag	ii
Figurer	iv
Tabeller	iv
3 Innledning	1
3.1 Motivasjon	1
3.2 Problemstilling	1
4 Teoretisk rammeverk og litteraturgjennomgang	1
4.1 Sammenheng mellom sosial klasse og elevprestasjoner i tidligere litteratur	2
4.2 Skoleproduktfunksjonen	3
4.3 Den politiske og økonomiske situasjonen i Brasil og Norge	3
5 Datamaterialet	6
5.1 Utførelse	6
5.2 Beskrivelse av den avhengige variabelen	6
5.3 Beskrivelse av de sentrale uavhengige variablene	7
5.3.1 Utdanning	7
5.3.2 Jobb	8
5.3.3 Bøker i hjemmet	8
5.3.4 Kjønn	8
5.3.5 Fødeland	8
5.3.6 Fravær	8
5.3.7 Språk hjemme	8
5.4 Deskriptiv statistikk for avhengig variabel	9
5.5 Deskriptiv statistikk for uavhengige variabler	9
5.6 Kritikk av datasettet	9
6 Empirisk strategi	10
6.1 Valg av variabler	10
6.2 Estimeringsmetode	10
6.3 Determinasjonskoeffisienten	11
6.4 Hypotesetesting	11

6.4.1	T-test	11
6.4.2	F-test	12
7	Økonometrisk modell	13
7.1	Funksjonsform	13
8	Empiriske resultater	15
8.1	Resultater fra regresjonsanalyse for Norge	15
8.2	Resultater fra regresjonsanalyse for Brasil	18
8.3	Sammenligning	21
9	Diskusjon	21
9.1	Videre arbeid	22
10	Oppsummering og konklusjon	22
A	R^2 for bruk i F-test	23
	Bibliografi	24

Figurer

1	«The Great Gatsby Curve». Den horisontale akse viser landets Gini indeks, og den vertikale akse viser intergenerasjonell elastisitet av inntekt.	4
2	Sammenligning av Norges og Brasils score på fem av Geert Hofstedes seks kulturelle dimensjoner; makt-distanse, individualisme, usikkerhetsaversjon, langtidsorientering, og indulgens. Data hentet fra [27].	5

Tabeller

1	Indeks for stabilitet og fravær av vold/terrorisme for Norge og Brasil fra 2018 til 2022.	4
2	Variabler brukt i vår analyse, og beskrivelse av dem.	7
3	Deskriptiv statistikk for den avhengige variabelen, <i>lese</i> , i Norge og Brasil.	9
4	Deskriptiv statistikk for de uavhengige variablene i analysen, i Norge og Brasil.	9
5	Resultat fra regresjonsanalyse for Norge. * $p < 0.05$, t-score i parentes	16
6	Resultat fra t-test av resultatene for Norge, ved et 5% signifikansnivå.	17
7	Resultat fra F-test av resultatene for Norge, ved et 5% signifikansnivå.	17
8	Resultat fra regresjonsanalyse for Brasil. * $p < 0.05$, t-score i parentes.	19
9	Resultat fra t-test av resultatene for Brasil, ved et 5% signifikansnivå.	20
10	Resultat fra F-test av resultatene for Brasil, ved et 5% signifikansnivå.	20

11	Determinasjonskoeffisienten R^2 fra regresjon utført uten restriksjoner for modellene B og C, til bruk i F-test for Norge.	23
12	Determinasjonskoeffisienten R^2 fra regresjon utført uten restriksjoner for modellene B og C, til bruk i F-test for Brasil.	23

3 Innledning

3.1 Motivasjon

«Det er bedre å bli født rik enn å bli født smart» påstår Anthony P. Carnevale i boken «The Merit Myth: How our colleges favor the rich and divide America» [1]. Carnevale, som er professor og leder for Georgetown University Center on Education and the Workforce i Washington, retter kommentaren mot rollen høyere utdanning spiller i et USA der de økonomiske ulikheten øker. Allerede i 1966 publiserte U.S. Department of Health, Education and Welfare en banebrytende rapport kalt «Equality of Educational Opportunity», som utforsket forholdet mellom sosial klasse og elevprestasjoner [2]. Rapporten, som ble etterspurt av borgerrettighetsloven fra 1964, konkluderte med at det var tydelige sammenhenger mellom sosioøkonomisk- og familiær bakgrunn og elevprestasjoner i USA på 60-tallet. Forfatteren, James Coleman, konkluderte med at afro-amerikanske elever fra husholdninger med lavere inntekt gjorde det betydelig dårligere på skolen sammenlignet med hvite elever fra rikere familier. Dette har vært fundamentet for lignende forskning i senere år. I Norge har nyere forskning også vist en sammenheng mellom sosial klasse og elevprestasjoner, noe som tyder på at dette er et globalt fenomen [3].

3.2 Problemstilling

Vi har valgt følgende problemstilling for vår bacheloroppgave:

«Hvordan avhenger elevprestasjoner i Norge og Brasil av sosial klasse?».

Vi kommer til å arbeide med problemstillingen med utgangspunkt i tall fra PIRLS-undersøkelsen som undersøker leseferdigheter hos 10-åringer i begge land. Vi kommer til å jobbe ut ifra hypotesen om at de negative effektene av lav sosial klasse på elevprestasjoner vil være større i Brasil enn i Norge, ettersom de sosiale forskjellene er større i Brasil enn i Norge [4]–[7]. I tillegg vil vi bruke kjønn som en supplerende hovedforklaringsvariabel, ettersom forskning tilsier at de negative effektene av sosial klasse er større hos menn videre i livet [8]. Vi ønsker å undersøke om vi kan se tegn på dette allerede hos 10-åringer. Som kontrollvariabel bruker vi fødeland, fravær, og språk snakket hjemme, da dette er faktorer som er vist å ha stor innvirkning på elevprestasjoner i grunnskolen [9], [10].

4 Teoretisk rammeverk og litteraturgjennomgang

I denne delen av oppgaven er målet å presentere tidligere litteratur som har sett på sammenhengen mellom skoleprestasjoner og sosial klasse. Skoleproduktfunksjonen vil også bli presentert, og denne vil være et fundament for den videre utvidelse av den økonomiske modellen presentert i Kapittel 7. Til slutt vil vi ta en nærmere titt på den sosioøkonomiske og politiske situasjonen i Norge og Brasil for å presentere forskjeller. Sosial klasse er betraktet som en utslagsgivende parameter når det kommer til elevprestasjoner. Hvor mye dette påvirker er dog ikke lett å si. Mye av grunnlaget for dette er naturligvis at bakgrunnen for elevers prestasjoner er svært sammensatt, og det er derfor vanskelig å kunne isolert sett si hvor mye hver faktor påvirker. Det er rimelig å gå ut ifra at høyere sosial klasse gir en positiv effekt på elevprestasjoner. Dette skyldes at sosial klasse ofte korrelerer med blant annet tilgang på ressurser og følelse av psykologisk trygghet. Samtidig er det andre faktorer som kan spille en betydelig rolle i elevens prestasjonsevne - slik som individuelle talent, innsats og personlig motivasjon.

4.1 Sammenheng mellom sosial klasse og elevprestasjoner i tidligere litteratur

Som nevnt innledningsvis baserer så godt som all forskning på tema sosial klasse og elevprestasjoner seg på James Colemans rapport fra 1966 [2]. Som resultat av at det store flertallet av amerikanske barn gikk på segregerte skoler på 60-tallet var det lett å studere ulike elevgrupper hver for seg, noe som også var en del av bakgrunnen for undersøkelsen. Rapportens hovedfunn var at elever fra minoritetsbakgrunner jevnt over gjorde det dårligere på skolen enn hvite elever. I tillegg viste Coleman at familiær bakgrunn, med faktorer som foreldres utdanningsnivå og inntekt, samt tilgang på bøker og aviser i hjemmet, hadde positiv effekt på prestasjoner. Dette minner om det vi i vår oppgave omtaler som «*sosial klasse*», og dette rettferdiggjør også bruk av Colemans rapport som en relevant primærkilde i vårt arbeid med bacheloroppgaven.

Som nevnt har også Coleman-rapporten inspirert mer moderne forskning rundt temaet i andre land, hvor trenden ligner det som ble observert i 1966. Foreldres inntekt er kanskje et av de mest nærliggende målene på sosial klasse, og dette er naturligvis noe av det nyere forskning på emnet har fokusert mest på. Blant annet indikerer en Britisk studie fra 2022 at Colemans funn fortsatt er relevant i dag; barn fra lav sosioøkonomisk bakgrunn i Storbritannia gjør det dårligere på skolen enn barn fra mer privilegerte husholdninger, og denne sammenhengen har vært konstant de siste 95 årene [11].

Den samme korrelasjonen blir også påvist i en Tsjekkisk studie, men her argumenteres det for at sammenhengen i større grad er indirekte. Studien påviser en signifikant sammenheng mellom foreldres inntekt og barnas skoleprestasjoner, men viser også at foreldres utdanning spiller en signifikant rolle [12]. Dette vises både som en direkte og indirekte effekt på barnas prestasjoner, både ved at foreldre med høy utdanning i større grad kan hjelpe med skolearbeid grunnet kompetanse, men også ved at foreldre med høy utdanning ofte påvirker barnas aspirasjon og motivasjon for å gjøre det bra på skolen.

Forskning fra Australia i 2013 finner også i likhet med Coleman en sammenheng mellom sosial klasse og elevprestasjoner. I artikkelen «Why poor children are more likely to become poor readers: The school years» finner Jennifer Buckingham, Kevin Wheldall og Robyn Beaman-Wheldall evidens for at barn fra lav sosioøkonomisk bakgrunn presterer dårligere på skolen, særlig fordi effekten av tidlige leseferdigheter på videre akademiske prestasjoner observeres som stor [13]. Tilgangen på bøker i hjemmet beskrives som en helt sentral faktor når det kommer til leseferdigheter, ettersom dette naturligvis påvirker tidlige leseferdigheter. En rapport fra Australian Council for Educational Research (ACER) i 2012 bekrefter denne sammenhengen, der australske elever med mange (200 eller flere) bøker gjennomsnittlig scorer 64 poeng bedre enn elever med få (25 eller færre) bøker [14]. Dette reflekterer også generell sosioøkonomisk klasse ifølge rapporten.

I studien «Social Class and Test Performance» fra 2011 finner forfatterne også en tydelig korrelasjon mellom foreldres inntekt og elevers presentasjoner. Korrelasjonen viser at mer velstående elever fikk bedre testresultater, delvis på grunn av bedre læringsressurser og læringsmiljø [15]. Forfatterne fremlegger også noe de kaller en stereotypetrussel. Dette forklarer de som når et individ presterer dårligere på grunn av negative stereotyper tilknyttet deres sosiale klasse. Konseptet symbolsk vold, som Pierre Bourdieu utviklet, omtales i teksten. Det beskriver hvordan sosiale normer og symboler, som virker naturlige og uunngåelige, faktisk skjuler de underliggende maktforholdene som skaper dem. I konteksten av symbolsk vold innen utdanning omhandler seg i praksis, læreplaner og testing som favoriserer visse sosiale grupper over andre, basert på kriterier som presenteres som meritokratiske, men som faktisk er partiske. Forfatterne etterlyser og mener at man trenger en systemendring for å oppnå en ekte meritokrati i utdanningen.

Dersom vi tar for oss utdanningssektoren i Norge og Brasil ser vi at de samme trendene går igjen.

Her til lands er det stor enighet om at sosial bagrunn har stor betydning for elevers skoleprestasjoner [3]. I tillegg til å påvirke læringsresultater, vil sosial klasse også spille inn på elevens ønske om videre utdanning og fremtidig yrkesvalg. Anton Hoëm beskriver dette med sin sosialiseringsteori, som beskriver hvordan ulike sosialiseringmønstre i skolen og senere i yrkeslivet oppstår gjennom samspillet mellom hjemmemiljøet og skole- og yrkesmiljøet. I en artikkel fra 2022 analyserer en

forskningsgruppe fra Lemann Center for Brazilian Studies akademiske prestasjonsforskjeller over en tiårsperiode i Brasil [16]. Forfatterne konkluderer at det eksisterer store prestasjonsulikheter blant 5. klassinger, og at disse ulikhetene økte i løpet av perioden. Ulikhetene skyldtes delvis stor ulikhet mellom skoler med elever fra høyere og lavere gjennomsnittlig sosial klasse, og mellom kommuner med ulik inntekt per innbygger og forskjellige utdanningspolitikk.

4.2 Skoleproduktfunksjonen

Skoleproduktfunksjonen stammer fra sosiologen James Colemans rapport fra 1966, hvor han undersøkte muligheten for å redusere raseforskjeller i USA ved å forbedre utdanningen til den svarte befolkningen [17]. Coleman konkluderte med at variabler innenfor familie- og medelevkarakteristika hadde signifikante verdier. Effekten av skoleinnsatsfaktorer, som lærerkarakteristika og skoleressursvariabler, viste seg derimot å være uklare og preget av variabler med lite signifikans i sine parametre. Resultatet fra undersøkelsen ga opphav til uttrykket «it's all in the family».

I artikkelen «Utforming av utdanningspolitikken – Hva kan økonomene bidra med?» forklarer Hans Borønningen funksjonen på følgende måte [17]:

$$T = f(F, P, S). \quad (1)$$

Variablene representerer følgende faktorer:

- T = elevprestasjoner,
- F = familie- og elevkarakteristika,
- P = medelevkarakteristika,
- S = skoleinnsatsfaktorer.

Elevprestasjoner varierer altså som en funksjon av de tre faktorene F, P og S. I vår oppgave bruker vi variablene som representerer familie- og elevkarakteristika; foreldrenes yrke og utdanning, antall bøker i hjemmet, kjønn, landet eleven er født i, fravær, og språket familien prater i hjemmet.

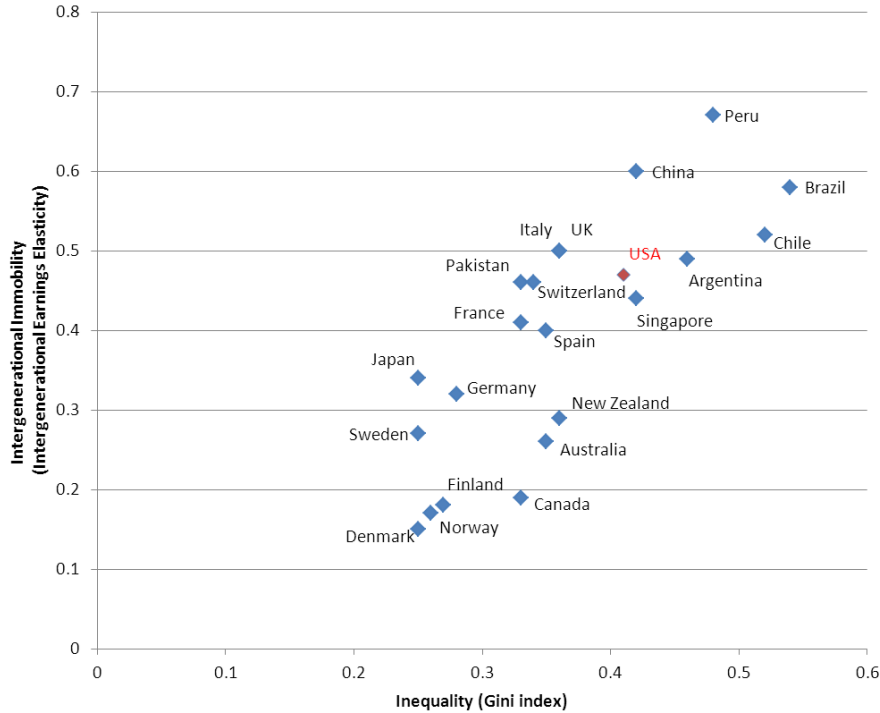
4.3 Den politiske og økonomiske situasjonen i Brasil og Norge

Vi har valgt å begrense analysen vår til to land: Norge og Brasil. Dette skyldes primært de sosioøkonomiske, politiske og kulturelle forskjellene mellom de to landene.

Brasil er en av verdens største økonomier, med en BNP på 1.92 billioner USD, men landet har gjennom historien hatt betydelige økonomiske problemer [4]–[6]. Per 2018 hadde Brasil det nest høyeste nivået av ulikhet blant OECD- og partnerland [18]. De 40% fattigste av inntektsmottakerne mottar kun 10% av disponible inntekter, mens de rikeste 10% mottar mer enn 14 ganger så mye. Brasils økonomi har vært preget av høy inflasjon, finansielle kriser, og stor økonomisk ulikhet. Historisk har lave skatteinntekter og høye offentlige utgifter ført til vedvarende offentlig underskudd, som har blitt finansiert gjennom lån. Norge har derimot en mindre økonomi, med BNP på 593.35 milliarder USD. Til tross for at ulikheten også har vokst her til lands de siste årene, er de økonomiske ulikhetene fortsatt relativt små [7]. Dette har historisk sikret stabil og høy økonomisk vekst. Den store forskjellen i BNP mellom landene kan også forklares ved forskjellen i befolkning; med et innbyggertall på 215.3 millioner er Brasil verdens sjetteste største land, mens det i Norge kun bor en brøkdel av dette; 5.457 millioner [5].

«The Great Gatsby Curve» er basert på forskningen til Miles Corak, og ble popularisert etter en tale av Alen Krueger i 2012. Begrepet omhandler det positive empiriske forholdet mellom inntektsulikhet og vedvarende inntekt over generasjoner. Kurven illustrerer et mønster der land med lav ulikhet ofte har høy sosial mobilitet og omvendt. I Figur 1 ser vi at Norge og Brasil er på

motsatt ende av denne kurven. Dette vil tilsi at man kan forvente at elevpresentasjonene er sterkere avhengig av sosial klasse i Brasil enn Norge. Den horisontale akse viser ulikhet, målt med landets Gini-indeks. Gini-indeksen viser inntektsforskjeller summert for hele befolkningen på en skala fra 0, som tilsvarer fullstendig likhet, til 1, som tilsvarer fullstendig ulikhet [19]. Den vertikale akse viser intergenerasjonell elastisitet av inntekt (IGE), som vil si at et land med høy sosial mobilitet vil ha en lav IGE.



Figur 1: «The Great Gatsby Curve». Den horisontale akse viser landets Gini indeks, og den vertikale akse viser intergenerasjonell elastisitet av inntekt.

Norge har et velutviklet velferdssystem, med gratis utdanning, et godt helsevesen og gode sosial- og trygdeordninger [20], [21]. Utdanningen i Norge er basert på prinsippet om likeverdighet og tilpasset opplæring for alle elever, og innebærer obligatorisk utdanning i ti år. Brasil har også etablert offentlig finansiert utdannings- og helsesystem, med obligatorisk utdanning for barn i alderen 7 til 14 år [2], [22]. Imidlertid er tilgangen på kvalitetsutdanning og helsetjenester sterkt varierende avhengig av geografisk beliggenhet og sosioøkonomisk status [23].

Norge både er og har historisk vært preget av politisk stabilitet, med en sterk sosialdemokratisk tradisjon [20]. Brasil har historisk vært preget av politisk uro, og store politiske omveltninger [24]. I dag er det politiske klimaet i Brasil preget av polarisering. Kvantitativt kan vi sammenligne situasjonen i de to landene ved å se på indeks for stabilitet og fravær av vold/terrorisme [5]. Indeksen vurderer oppfatninger om sannsynligheten for politisk ustabilitet og/eller politisk motivert vold, inkludert terrorisme. Estimater angir landets poengsum på den aggregerte indikatoren, målt i enheter av en standard normalfordeling, dvs. med en spennvidde fra omtrent -2.5 til 2.5, hvor -2.5 tilsvarer høy og 2.5 tilsvarer lav grad av politisk ustabilitet og politisk motivert vold. Verdiene for Norge og Brasil vises i Tabell 1, fra 2018 til 2022.

Land	2018	2019	2020	2021	2022
Norge	1,12	1,14	1,22	1,09	0,86
Brasil	-0,46	-0,71	-0,45	-0,42	-0,33

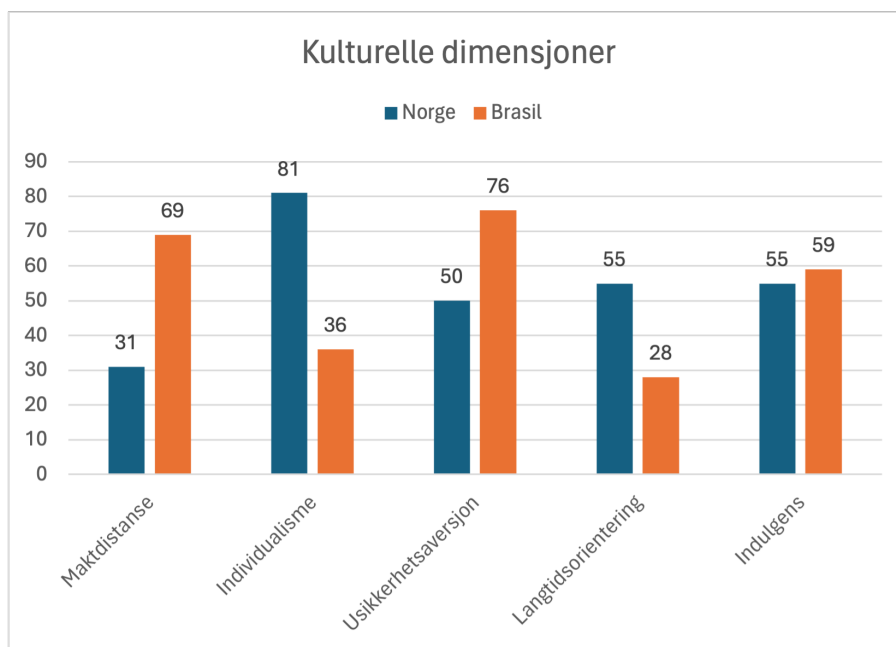
Tabell 1: Indeks for stabilitet og fravær av vold/terrorisme for Norge og Brasil fra 2018 til 2022.

Som vi ser fra Tabell 1, har Norge konsekvent scoret høyere på indeksen enn Brasil, som tilsvarer en mer politisk stabil situasjon i landet. Derimot ser vi også at situasjonen har forbedret seg i Brasil de siste årene, mens Norge har blitt mer politisk ustabil.

For å se på de kulturelle forskjellene mellom de to landene skal vi benytte Geert Hofstedes teori om kulturelle dimensjoner [25]. Dimensjonene reflekterer ulike holdninger og verdier, og er preget av nasjonal tilhørighet [26]. I denne oppgaven vil vi primært ta for oss fem av de totalt seks kulturelle dimensjonene, som er som følger:

- stor vs. liten maktdistanse
- kollektivism vs. individualisme
- unngår vs. takler usikkerhet
- langtids- vs. korttidsorientering
- mye vs. lite indulgens.

Figur 2 viser hvordan Norge og Brasil scorer på fem av disse dimensjonene.



Figur 2: Sammenligning av Norges og Brasils score på fem av Geert Hofstedes seks kulturelle dimensjoner; maktdistanse, individualisme, usikkerhetsaversjon, langtidsorientering, og indulgens. Data hentet fra [27].

Når vi sammenligner Norge og Brasil ser vi at landene ligger på motsatte ender av maktdistanse, individualisme, usikkerhetsaversjon, og langtidsorientering. Kulturelle dimensjoner gjenspeiles ofte i landets utdanningssystem; for eksempel er det vanlig at kollektivistiske kulturer, slik som Brasil, fokuserer på hele gruppen i undervisningen, og lar elevene lære fra hverandre [28]. Individualistiske kulturer, slik som Norge, tenderer mot å fokusere på enkeltindivider i undervisningen. I kulturer med lite maktdistanse opplever elever økt tilhørighet til skolen sin [29]. Tilhørighet til utdanningsinstitusjonen igjen er assosiert med elevens oppfattet støtte fra læreren og en preferanse for samarbeidskultur i klasserommet. Kulturer preget av langtidsorientering gir ofte bedre forutsetninger for innvandrere i utdanningsinstitusjonene [30]. Her presterer elever med innvanderbakgrunn bedre i lese- og regnetester på 3. klasse nivå, de har bedre prestasjonsøkning over tid, mindre fravær, samt høyere sannsynlighet for å fullføre videregående innen normert tid. Derfor er det interessant å se om vi får resultater som gjenspeiler dette, spesielt når vi ser på prestasjoner hos elever som er født i et annet land, eller snakker annet språk i hjemmet.

5 Datamaterialet

Datasettet som er brukt i denne oppgaven er hentet fra den internasjonale leseundersøkelsen Progress in International Reading Literacy Study (PIRLS) [31]. PIRLS er en internasjonal leseundersøkelse som utføres hvert femte år av International Association for the Evaluation of Educational Achievement (IEA). Undersøkelsen har som mål å følge utviklingen av leseferdighetene til 10-åringer over tid. Datasettet som vi har brukt er den nyeste undersøkelsen, gjort i 2021, hvor 66 land deltok, deriblant Norge og Brasil. PIRLS ble for første gang gjennomført i 2001, men 2021 var første året Brasil deltok.

5.1 Utførelse

PIRLS er en utvalgsundersøkelse, og i 2021 var det 3,589 og 4,941 som deltok fra henholdsvis Norge og Brasil [31], [32]. Internasjonalt er prøven lagt til slutten av 4. trinn, mens Norge utfører testen på elever i 5. trinn for å lettere kunne sammenligne seg med andre nordiske land, som har førskoleklasser. Undersøkelsen utføres ved at elevene mottar et oppgavesett som inneholder to tekster, eksempelvis en litterær tekst og en informasjonstekst. Elevene skal lese og besvare spørsmål knyttet til tekstene.

I tillegg til å kartlegge hvor godt elevene leser, kartlegger PIRLS også viktig bakgrunnsinformasjon fra spørreskjema. Foreldre og elever mottar spørreskjema som skal måle ulike elev- og familiekarakteristika, som for eksempel alder, kjønn, innvandrerstatus, fravær, osv. Lærerne og skolens ledere mottar spørreskjema rettet mot lærerkarakteristika, som for eksempel klassestørrelse, alder, kjønn, utdanningsnivå, osv., ressurskarakteristika, for eksempel tilgang på PC, og skolekarakteristika, for eksempel prosentandelen av elevene på skolen fra vanskeligstilte hjem, prosentandelen av elevene på skolen som ikke er født i landet, antall innbyggere i området hvor skolen er lokalisert, og så videre.

5.2 Beskrivelse av den avhengige variabelen

Formålet med de påfølgende avsnittene er å gå dypere inn i både hvilke variabler vi har valgt, og hvordan de fungerer i denne analysen. Det er viktig å presisere at analysen vil sammenligne to regresjonsmodeller, en for hvert land, og hver variabel vil da naturligvis eksistere i to eksemplarer. Variablene er også presentert i Tabell 2.

Variabel	Beskrivelse
<i>leseferdigheter</i>	Gjennomsnittlig poengsum på lesetesten til PIRLS blant deltakere
<i>høyere_utdanning</i>	1 = Om foreldrene har gjennomført årsstudium, bachelor eller master 0 = Ellers
<i>profesjonell_jobb</i>	1 = Om foreldrene jobber som kontoransatt, håndverker, bedriftsleder, profesjonell, eller assisterende profesjonell 0 = Om foreldrene jobber som liten bedriftseier, jordbruk/fiske-arbeider, salg/service-arbeider, maskinoperatør, allmennarbeider, eller aldri har jobbet
<i>mange_bøker</i>	1 = Mer enn 25 bøker i hjemmet 0 = 25 eller færre bøker i hjemmet
<i>mange_barnebøker</i>	1 = Mer enn 25 barnebøker i hjemmet 0 = 25 eller færre barnebøker i hjemmet
<i>gutt</i>	1 = Gutt 0 = Jente
<i>ikke_født</i>	1 = Eleven er ikke født i landet 0 = Eleven er født i landet
<i>mye_fravær</i>	1 = Eleven har fravær en gang i uka, eller en gang annenhver uke 0 = Eleven har fravær sjeldnere
<i>språk_hjemme</i>	1 = Eleven snakker testens språk hjemme alltid eller nesten alltid 0 = Eleven snakker testen språk hjemme sjeldnere

Tabell 2: Variabler brukt i vår analyse, og beskrivelse av dem.

Vi benytter *leseferdigheter* som avhengig variabel i vår analyse. For enkelhets skyld, og med hensyn til senere økonometrisk modell, kaller vi denne variabelen enkelt og greit for *lese*. Denne beskriver altså gjennomsnittlig poengsum på lesetesten til PIRLS blant deltagerne fra Norge og Brasil. Det er verdt å nevne at denne variabelen består av data fra kun 1 av 5 plausible verdi-variabler på testscore, som svekker analysen i forhold til usikkerhet, robusthet og representativitet. Derimot vil vi hevde at en mer omfattende analyse på alle fem variabler vil medføre en langt mer komplisert prosess enn det som er hensiktsmessig ved en bacheloroppgave, som rettfærdiggjør forenklingen vi velger å gjøre.

5.3 Beskrivelse av de sentrale uavhengige variablene

De fire mest sentrale forklaringsvariablene som utgjør vår definisjon på sosial klasse i analysen er dummykodet, slik at de tar verdi 0 eller 1. Bruken av dummyvariabler øker nøyaktighet i modellen, ettersom en kategorisk variabel, uten like intervaller mellom kategoriene, ikke egner seg for å bli håndtert kvantitativt. I tillegg legger vi til tre kontrollvariabler for å isolere effekten av hovedforklaringsvariablene våre: om eleven er født i landet den tar testen i, hvor mye fravær eleven har og om eleven snakker testens språk hjemme. Vi mener det er sannsynlig at disse variablene også kan påvirke testresultatene til elevene, og ved å kontrollere for disse reduserer vi risikoen for at svarene vi får fra analysen av hovedforklaringsvariablene våre blir forvridd, som dermed styrker analysens pålitelighet og validitet [9], [10].

5.3.1 Utdanning

Variabelen for høyeste fullførte utdanningsnivå hos foreldrene er omgjort fra International Standard Classification of Education (ICED) nivå 1 til 8, til å kun se på hvorvidt foreldrene har høyere utdanning eller ikke. Den omkodede variabelen, kalt *høyere_utdanning* er kodet slik at kategoriene «årsstudium», «bachelor», «master» og «PhD» utgjør 1 og alt annet er kodet lik 0. Dermed har vi omgjort en variabel som kategoriserer ulike grader av utdanning hos testdeltakernes foreldre, til en variabel som viser om foreldrene har høyere utdanning eller ikke. Vi mener at dette er den mest interessante dimensjonen for å studere effekten av utdanning på barns skoleprestasjoner. I tillegg er det verdt å nevne at kategorien «not applicable» er fjernet.

5.3.2 Jobb

Foreldrenes jobb uttrykkes gjennom dummyvariabelen *profesjonell_jobb*, der kategoriene er fordelt på følgende måte:

Ikke-profesjonell jobb:

- Aldri jobbet
- Liten bedriftseier
- Jordbruk/fiske-arbeider
- Salg/service-arbeider
- Maskinoperatør
- Allmennarbeider

Profesjonell jobb:

- Kontoransatt
- Håndverker
- Bedriftsleder
- Profesjonell
- Assisterende profesjonell

Her er kategoriene som er plassert innenfor begrepet ikke-profesjonell jobb ment å reflektere jobber som krever mindre utdanning eller spesialisering, og motsatt for jobbene under kategorien profesjonell jobb. Dette er ment å skille såkalte lav-klasse og høy-klasse jobber for å se effekten av dette på den avhengige variabelen. Ikke-profesjonell jobb er kodet som 0 profesjonell jobb er kodet som 1. Det er også viktig å nevne at kategorien «not applicable» er fjernet fra variabelen. Merk at *profesjonell_jobb* måler den høyeste nivået av ansattforhold blant foreldrene.

5.3.3 Bøker i hjemmet

Antall bøker i hjemmet er kodet slik at $= < 25$ bøker = 0, angitt som få bøker, og > 25 bøker = 1, angitt som mange bøker. Dette gir variabelen *mange_bøker*. Kodingen er helt lik for variabelen for barnebøker, som i vår oppgave kalles *mange_barnebøker*.

5.3.4 Kjønn

Kjønn er også dummykodet, slik at 0 = jente og 1 = gutt, som dermed gir opphav til variabelen *gutt*. For Brasil-settet er kategorien «other» fjernet.

5.3.5 Fødeland

Første kontrollvariabel, kalt *ikke_født*, er dummykodet slik at elever som er født i landet de tar testen i, henholdsvis Norge eller Brasil, faller under kategori 0, og elever ikke født i landet går under kategori 1.

5.3.6 Fravær

Variabelen *mye_fravær* er kodet slik at kategoriene «fravær en gang i uka» og «fravær en gang annenhver uke» = 1, og går under den nye kategorien *mye_fravær*. Fravær sjeldnere enn dette er kodet lik 0, og omtales som lite fravær.

5.3.7 Språk hjemme

I variabelen *språk_hjemme* har vi kodet det slik at dersom eleven snakker testens språk hjemme «alltid» eller «nesten alltid» er dette gitt som 1, og noe sjeldnere enn dette er gitt lik 0.

5.4 Deskriptiv statistikk for avhengig variabel

I Tabell 3 presenterer vi deskriptiv statistikk for den avhengige variabelen, *lese*. Vi ser på variabelen både i Norge og Brasil.

Tabell 3: Deskriptiv statistikk for den avhengige variabelen, *lese*, i Norge og Brasil.

	Norge	Brasil
Gjennomsnitt	553.3679	435.3794
Min	268.8082	19.13124
Max	777.7022	798.1216
Standardavvik	69.95707	118.104
Antall observasjoner	3589	4494

Gjennomsnittlig testresultat er cirka 118 poeng høyere i Norge enn i Brasil. Standardavviket er høyere i Brasil, som viser spredningen i elevprestasjoner er større der. Vel å merke er det 905 flere elever som deltok på undersøkelsen i Brasil enn i Norge.

5.5 Deskriptiv statistikk for uavhengige variabler

I Tabell 4 presenterer vi deskriptiv statistikk for de uavhengige variablene. Vi ser på variablene både i Norge og Brasil.

Tabell 4: Deskriptiv statistikk for de uavhengige variablene i analysen, i Norge og Brasil.

	Norge		Brasil	
	Gjennomsnitt	Standardavvik	Gjennomsnitt	Standardavvik
<i>høyere_utdanning</i>	0.85	0.36	0.34	0.47
<i>profesjonell_jobb</i>	0.88	0.32	0.38	0.49
<i>mange_bøker</i>	0.84	0.37	0.19	0.39
<i>mange_barnebøker</i>	0.83	0.37	0.16	0.37
<i>gutt</i>	0.51	0.50	0.51	0.50
<i>ikke_født</i>	0.06	0.23	0.01	0.12
<i>mye_fravær</i>	0.06	0.24	0.03	0.44
<i>språk_hjemme</i>	0.94	0.23	0.99	0.11

I Norge er det en betydelig høyere andel av foreldre som har høyere utdanning og profesjonell jobb. Standardavviket for høyere utdanning og arbeidsforhold er større for Brasil enn for Norge, som tilsier at spredningen er større der. Norge har en litt høyere andel elever som ikke er født i landet, og naturlig nok er det derfor også en litt høyere andel som ikke snakker testens språk i hjemmet i Norge. Norge har også en litt høyere andel elever med mye fravær.

5.6 Kritikk av datasettet

Som nevnt i 4.1 understreker Coleman betydningen av familiekarakteristika for elevprestasjonene. En viktig parameter innen familiekarakteristika, spesielt for å måle den sosioøkonomiske situasjonen, er inntekt. Datasettet fra 2021 mangler dette, og vil dermed ikke gi et fullstendig bilde av familiens sosioøkonomiske situasjon. Derfor har vi valgt å inkludere en dummyvariabel for jobb, hvor vi skiller mellom profesjonell og ikke-profesjonell jobb. Det er verdt å merke seg at denne tilnærmingen ikke nødvendigvis er den mest nøyaktige eller perfekte proxyen for familiens økonomiske status, da det finnes ikke-profesjonelle yrker som kan tjene mer enn profesjonelle. Likevel

mener vi at det er en akseptabel proxy, da det generelt gir en indikasjon på familiens økonomiske situasjon og kan bidra til å kompensere for manglende inntektsdata.

Analysen vår er begrenset til observasjoner av elever i 4. og 5. klasse for henholdsvis Brasil og Norge. Ideelt skulle datasettet inkludert observasjoner gjort av elever i det samme klassetrinnet, for å kontrollere for at ulikhetene i leseferdigheter ikke skyldes at elevene i Brasil har kommet kortere i utdanningsløpet. Det hadde også vært spennende å se hvordan effekten av kjønn endret seg over tid, da som nevnt de negative effektene av sosial klasse er større hos menn senere i livet.

Avslutningsvis består datasettet kun av 3589 og 4494 observasjoner for henholdsvis Norge og Brasil. Dette er et veldig lite utvalg, spesielt for store land som Brasil. For å kunne gjenspeile effekten av sosial klasse i den virkelige verden vil det være behov for flere observasjoner.

6 Empirisk strategi

6.1 Valg av variabler

Vår definisjon på sosial klasse eller sosioøkonomisk status er naturligvis basert på hva tidligere forskning benytter, som presentert i del 4.1, der faktorer som inntekt, utdanning og bøker i hjemmet går igjen i tidligere studier. Ved å benytte variabler som tidligere forskning også har brukt styrker dette blant annet sammenlignbarheten og troverdigheten i vår analyse. Våre resultater vil altså passe bedre inn i det eksisterende kunnskapslandskapet ettersom de stammer fra analyse av lignende variabler, og ved å bruke et anerkjent mål på sosial klasse styrker det resultatenes troverdighet i og med at metodologien som benyttes har vært i bruk siden 60-tallet. I tillegg vil bruk av en slik anerkjent metodologi styrke oppgavens validitet, ettersom det øker sannsynligheten for at vi måler det vi faktisk ønsker å måle.

Grunnlaget for vår definisjon av sosial klasse er i stor grad påvirket av PIRLS sin egen skala. Nytt for 2021-versjonen av undersøkelsen er det de kaller «Home Socioeconomic Status scale», der dataen som utgjør skalaen utelukkende stammer fra elevens foreldre. PIRLS bruker fire indikatorer i sin måling: antall bøker i hjemmet (unntatt barnebøker), antall barnebøker i hjemmet, foreldres utdanning og foreldres jobb [33]. Dette ligger i datasettet som en egen sammenslått variabel, men vi vil argumentere for at det er mer hensiktsmessig for vår analyse å skille ut hver indikator som egne variabler for å kunne observere og skille mellom effektene av hver enkelt del av PIRLS sitt mål på sosioøkonomisk status.

Vi mener i likhet med PIRLS at bruken av disse fire uavhengige variablene til en viss grad rettferdiggjør utelatelsen av «foreldres inntekt» som variabel, ettersom alle disse kan sies å være proxyer for økonomisk kapital. Derimot, som diskutert tidligere, erkjenner vi at det å finne en perfekt proxy for inntekt i praksis er veldig vanskelig, og en inntektsvariabel ville sannsynligvis vært hensiktsmessig å inkludere i tillegg, dersom PIRLS fortsatt hadde målt dette.

Som nevnt innledningsvis i oppgaven ønsker vi også å ta høyde for kjønn, der hypotesen er at lavere sosioøkonomisk status vil ha en større påvirkning på gutter enn jenter i barneskolealder, ettersom effekten av sosial klasse observeres som større hos menn enn kvinner i voksen alder [8]. Dermed har vi 5 uavhengige variabler i analysen som måler sosial klasse: kjønn, antall bøker i hjemmet (unntatt barnebøker), antall barnebøker i hjemmet, foreldres utdanning og foreldres jobb.

6.2 Estimeringsmetode

Oppgaven har som mål å undersøke sammenhengen mellom sosial klasse og elevprestasjoner - for å oppnå dette skal vi benytte oss av Minste Kvadraters Metode (OLS) [34]. OLS innebærer å minimere summen av kvadrerte avvik, eller residualer, som oppstår ved å sammenligne de observerte verdiene med de predikerte verdiene. Ved bruk av OLS observerer vi en lineær sammenheng mellom forklaringsvariablene og en avhengig variabel i en modell. Denne sammenhengen visualiseres gjennom en regresjonslinje, som illustrerer hvordan den avhengige variabelen endrer seg når den

uavhengige variabelen varierer.

Dersom vi betrakter en enkel modell:

$$y = \alpha + \beta x$$

der β er stigningstallet, og α er et konstantledd/ skjæringspunkt. Da vil den estimerte regresjonslinjen kunne beskrives som følger:

$$\hat{y} = \hat{\alpha} + \hat{\beta}\hat{x}$$

En enkel regresjonslikning kan uttrykkes som: $Y = \alpha + \beta_1x_1 + \dots + \beta_ix_i + \epsilon$, der ϵ er et restledd. OLS krever følgende forutsetninger om restleddet:

1. Forventningen til restleddet er lik null.
2. Variansen til restleddet er konstant.
3. Det er ingen korrelasjon mellom restleddene.
4. Restleddet er normalfordelt.

6.3 Determinasjonskoeffisienten

Determinasjonskoeffisienten, R^2 , indikerer hvor mye av variasjonen i den avhengige variabelen som kan forklares av de uavhengige variablene i modellen [34]. R^2 har en verdi mellom 0 og 1. En verdi nær 0 antyder ingen korrelasjon, mens en verdi nær 1 indikerer at modellen forklarer mesteparten av variansen i den avhengige variabelen.

Determinasjonskoeffisienten kan uttrykkes på følgende måte:

$$R^2 = 1 - \frac{SSR}{SST}$$

Her representerer SSR den forklarte variansen og SST den totale variansen i den avhengige variabelen.

6.4 Hypotesetesting

For å teste de tidligere formulerte hypotesene om sammenheng mellom leseferdigheter og sosial klasse benytter vi to ulike typer hypotesetester, som med tall fra regresjonsresultatene forteller om eventuelle funn er statistisk signifikante [34]. For å begynne en slik test må vi først formulere statistiske hypoteser om populasjonen; både en nullhypotese, H_0 , og en alternativ hypotese, H_1 . Nullhypotesen uttrykker at det ikke finnes en sammenheng mellom avhengig og uavhengig variabel, og dersom man ønsker å identifisere en signifikant sammenheng mellom variablene blir målet dermed naturligvis å finne grunnlag for å forkaste nullhypotesen. I vår analyse vil vi kjøre både t-tester og f-tester for å undersøke henholdsvis isolerte sammenhenger mellom enkelte uavhengige variabler og leseferdigheter, og hvordan modellene som helhet endrer forklaringskraft når flere variabler introduseres. Testene baserer seg på forutsetningene i 6.2 om at restleddet er normalfordelt, som igjen medfører at estimatorene er normalfordelt.

I denne oppgaven kommer vi til å bruke Stata og Python til å utføre hypotesetesting.

6.4.1 T-test

Generelt benytter vi en t-test for å teste statistiske hypoteser om et gjennomsnitt i populasjonen med utvalgsgjennomsnittet som grunnlag [34]. En testobservator for en slik test er vanligvis skrevet som:

$$t = \frac{(\text{estimat} - \text{hypotetisk verdi})}{\text{standardfeil}} \quad (2)$$

Her refererer «estimat» til regresjonskoeffisienten vi henter fra Stata-analysen. Hypotetisk verdi er generelt lik 0 i t-tester, ettersom vi tester direkte mot ingen effekt [35]. Standardfeil for et gjennomsnitt er gitt som standardavviket på kvadratroten av antall observasjoner i utvalget [34]. I oppgaven vil vi heretter referere til ligningen over med følgende notasjon:

$$TS = \frac{(b_i - \beta_i)}{sb_i} \quad (3)$$

I en t-test måler man resultat opp mot en kritisk verdi for å kunne indikere statistisk signifikans, som avhenger av grader av frihet og utvalgsstørrelse. I denne oppgavens tilfelle vil vi derimot argumentere for at utvalget er mer enn stort nok til å kunne anta normalfordeling, både i datasettet for Norge og Brasil. Med et 5% signifikansnivå kan vi derfor fastsette den kritiske verdien til 1.96, ifølge en standard normalfordelingstabell.

6.4.2 F-test

For å teste en multippel hypotese om flere parametere samtidig kan F-test benyttes. For eksempel kan en hypotese om at både β_1 og β_2 er lik 0 testes [34]. Ideen bak denne testen er å basere den på endring i residualkvadratsummen ($SSR = \sum e^2$) gitt restriksjoner. En betydelig endring i SSR tyder på at restriksjonene under nullhypotesen er ugyldige. Testobservatoren blir da F-fordelt, gitt signifikansnivå α :

$$T_S = \frac{(SSRR - SSRU)/h}{SSRU/(n - k)} \sim F_\alpha(h, n - k) \quad (4)$$

Her er:

- $SSRR$ - residualkvadratsummen med restriksjoner
- $SSRU$ - residualkvadratsummen uten restriksjoner
- h - antall restriksjoner under nullhypotesen
- k - antall parametere i modellen
- n - antall observasjoner

Ligning 4 kan også skrives om slik at man benytter seg av determinasjonskoeffisienten R^2 :

$$T_S = \frac{(R_U^2 - R_R^2)/h}{(1 - R_U^2)/(n - k)} \sim F_\alpha(h, n - k) \quad (5)$$

Her er:

- R_R^2 - determinasjonskoeffisienten med restriksjoner
- R_U^2 - determinasjonskoeffisienten uten restriksjoner
- h - antall restriksjoner under nullhypotesen
- k - antall parametere i modellen
- n - antall observasjoner

Beslutningsregelen blir da å forkaste H_0 dersom $T_S > F_\alpha(h, n - k)$.

Forskjellig fra t-testen, som studerer hvorvidt sammenhengene mellom variabler i en regresjonsmodell er statistisk signifikante, ser f-testen på hvorvidt en modell opplever en statistisk signifikant forbedring dersom flere variabler legges til. En slik test er naturligvis nyttig for vår analyse, etter som vi opererer med en basismodell som utvides to ganger.

7 Økonometrisk modell

Opgaven tar som nevnt utgangspunkt i James Colemans skoleproduktfunksjon, som tidligere formulert som ligning 1 i avsnitt 4.2, hvor denne fungerer som en basis til vår økonometriske modell.

7.1 Funksjonsform

Den metodiske kjernen og strukturen til oppgaven vår består av en lineær OLS-regresjonsmodell, ment til å beskrive sammenhengen mellom vår avhengige variabel, leseferdigheter, og våre uavhengige variabler. Vi har valgt en lineær modell fremfor en log-lineær modell av den grunn at estimatoren i den lineære modellen uttrykker endringer i nivået, ikke den relative endringen. Dette betyr at vi får en endring i testresultatets nivå per enhetsendring i variabelen. På den annen side ville en log-lineær modell gitt prosentvise endringer, noe som kan være mer utfordrende å tolke i denne analysen. Fra Colemans skoleproduktfunksjon er hovedfokuset i vår modell på det han omtaler som familie- og elevkarakteristika, som betyr at vi velger å se bort i fra medelevkarakteristika og skoleinnsatsfaktorer. Vi har valgt å ta utgangspunkt i en enkel basismodell, som er utledet med utgangspunkt i skoleproduktfunksjonen, presentert i avsnitt 4.2. Deretter utvider vi den to ganger, for å spisse analysen og styrke modellens nøyaktighet. Dermed har vi grunnlag for å utrede modellene A, B, og C.

Grunnmodellen A er gitt som følgende uttrykk:

$$\begin{aligned} lese = & \beta_0 + \beta_1 \text{profesjonell_jobb} + \beta_2 \text{høyere_utdanning} + \beta_3 \text{mange_bøker} \\ & + \beta_4 \text{mange_barnebøker} + \beta_5 \text{gutt} + \varepsilon_i \end{aligned} \quad (\text{A})$$

Grunnmodellen B er gitt som følgende uttrykk:

$$\begin{aligned} lese = & \beta_0 + \beta_1 \text{profesjonell_jobb} + \beta_2 \text{høyere_utdanning} + \beta_3 \text{mange_bøker} \\ & + \beta_4 \text{mange_barnebøker} + \beta_5 \text{gutt} + \beta_6 \text{profesjonell_jobb_gutt} \\ & + \beta_7 \text{høyere_utdanning_gutt} + \beta_8 \text{mange_bøker_gutt} \\ & + \beta_9 \text{mange_barnebøker_gutt} + \varepsilon_i \end{aligned} \quad (\text{B})$$

Grunnmodellen C er gitt som følgende uttrykk:

$$\begin{aligned} lese = & \beta_0 + \beta_1 \text{profesjonell_jobb} + \beta_2 \text{høyere_utdanning} + \beta_3 \text{mange_bøker} \\ & + \beta_4 \text{mange_barnebøker} + \beta_5 \text{gutt} + \beta_6 \text{profesjonell_jobb_gutt} \\ & + \beta_7 \text{høyere_utdanning_gutt} + \beta_8 \text{mange_bøker_gutt} + \beta_9 \text{mange_barnebøker_gutt} \\ & + \beta_{10} \text{ikke_født} + \beta_{11} \text{mye_fravær} + \beta_{12} \text{språk_hjemme} + \varepsilon_i \end{aligned} \quad (\text{C})$$

A: Den første modellen viser en tilnærmet estimering av effekten av arbeidsforhold og utdanningsnivå hos foreldrene, antall bøker og barnebøker i hjemmet, og kjønn i Norge og Brasil. Første ledd, β_0 , er konstantleddet som viser den gjennomsnittlige verdien til *leseferdigheter* når alle variablene er lik null. Videre utvider vi modellen i samsvar med skoleproduktfunksjonen.

B: Inkluderer interaksjonsledd for kjønn, dvs. verdien for arbeidsforhold, utdanningsnivå hos foreldrene, antall bøker i hjemmet, eller barnebøker i hjemmet multiplisert med verdien for kjønn.

Hensikten med et interaksjonsledd er å undersøke om forholdet mellom to variabler avhenger av verdien til en tredje variabel [36].

C: Kontrollerer for om kandidaten er født i landet, har mye fravær, og språk i hjemmet, som kombinerer både familie- og elevkarakteristika.

8 Empiriske resultater

I denne delen vil vi presentere resultatene fra OLS-regresjonsanalyse av modellene for både Norge og Brasil, samt supplere med hypotesetester av resultatene med bruk av teori fra del 4.2. Regresjonsresultatene for modell A, B, og C for Norge og Brasil er delt i to tabeller, henholdsvis Tabell 5 og 8. Vi vil først ta for oss en analyse av de norske resultatene.

8.1 Resultater fra regresjonsanalyse for Norge

Resultatene fra regresjonsanalysen for Norge er presentert i Tabell 5. Resultatene viser at alle variablene, med unntak av *gutt* (herunder også interaksjonsleddene som inkluderer kjønn) og *språk_hjemme* har en statistisk signifikant sammenheng ($p < 0.05$) med leseferdighetene til elevene. For eksempel har elever med foreldre i profesjonelle jobber, høyere utdanning, med tilgang til mange bøker og barnebøker i hjemmet en tendens til å ha bedre leseferdigheter. Dette indikerer at det vi omtaler som «sosial klasse» har en positiv effekt på leseferdighetene. Resultatene viser også at elever som ikke er født i Norge og elever med mye fravær, har dårligere leseferdigheter - dette er som forventet. Derimot er det noe uventet at effekten av å prate testens språk i hjemmet har en negativ effekt på elevens leseferdigheter, til tross for at denne effekten ikke er statistisk signifikant.

Videre observerer vi at gutter har betydelig lavere leseferdigheter enn jenter, men denne sammenhengen er kun signifikant i modell A. Interaksjonsleddene i modellen viser varierende effekter, noe som kan indikere at den positive effekten av sosioøkonomiske faktorer varierer avhengig av kjønn, til tross for at ingen av interaksjonsleddene er statistisk signifikante.

Introduksjonen av interaksjonsledd i modell B fører til redusert effekt av *høyere_utdanning* og *gutt*, mens effekten av *mange_barnebøker* holdt seg relativt konstant. Effekten av *profesjonell_jobb* og *mange_bøker* økte. Introduksjonen av kontrollvariabler i modell C hadde tilsvarende effekt relativt til A, dvs. at resultatene var tilsvarende som for B.

Samlet sett gir modellene en moderat forklaringsgrad på variasjonen i leseferdigheter, med et R^2 -verdi som varierer fra 0.0774 til 0.0928, avhengig av modell. Dette antyder at det er andre faktorer som også påvirker leseferdighetene til elevene, som ikke er inkludert i analysen.

Variabler	A <i>lese</i>	B <i>lese</i>	C <i>lese</i>
<i>profesjonell_jobb</i>	13.49185* (3.25)	18.74667* (3.10)	17.7309* (2.94)
<i>høyere_utdanning</i>	20.03993* (5.42)	15.42184* (2.82)	15.66265* (2.88)
<i>mange_bøker</i>	20.4712* (5.43)	23.89619* (4.43)	23.10854* (4.31)
<i>mange_barnebøker</i>	17.99267* (4.92)	17.3131* (3.24)	17.14148* (3.22)
<i>gutt</i>	-15.64433* (-6.70)	-9.539801 (-1.08)	-8.420286 (-0.96)
<i>profesjonell_jobb_gutt</i>		-9.643953 (-1.16)	-10.43192 (-1.26)
<i>høyere_utdanning_gutt</i>		8.159385 (1.10)	8.024524 (1.09)
<i>mange_bøker_gutt</i>		-6.476148 (-0.86)	-5.911482 (-0.79)
<i>mange_barnebøker_gutt</i>		1.222503 (0.17)	1.079335 (0.15)
<i>ikke_født</i>			-16.52572* (-2.80)
<i>mye_fravær</i>			-32.10472* (-6.67)
<i>språk_hjemme</i>			-6.997157 (-1.17)
konstant	502.7401* (109.73)	499.6844* (79.97)	510.3702* (62.24)
<i>N</i>	3,252	3,252	3,252
<i>R</i> ²	0.0774	0.0782	0.0928

Tabell 5: Resultat fra regresjonsanalyse for Norge.

* $p < 0.05$, t-score i parentes

For å sjekke om estimatorene foran interessevariablene er av signifikant betydning gjennomfører vi hypotesetester for hver av variablene. Vi benytter oss av teorien beskrevet i kapittel 6.4 og robuste standardavvik. For å teste enkeltparametre bruker vi t-test. Her ønsker vi å se på om estimatorene foran variablene er signifikant ved et 5% signifikansnivå. Her brukes en tosidig test, og derfor bruker vi $\alpha = \frac{5\%}{2} = 2.5\%$. Vi bruker Stata til å finne den kritiske verdien av t-distribusjonen ved å bruke kommandoen `invttail`, frihetsgrad tilsvarende utvalgsstørrelse N minus antall variabler K minus 1. Kritisk verdi, i kursiv, og resultatene fra t-testen er presentert i Tabell 6. Her representerer $>$ at absoluttverdien av TS er større enn kritisk verdi for den korresponderende estimatorene, som betyr at vi kan forkaste nullhypotesen til fordel for den alternative hypotesen. $<$ representerer at absoluttverdien av TS er mindre enn kritisk verdi, som betyr at vi ikke kan forkaste nullhypotese til fordel for den alternative hypotesen. Estimatorene tilsvarer slik de presenteres i A, B, og C. I alle tilfellene er nullhypotesen at verdien av estimatoren er lik null, og den alternative hypotesen er at $\beta_i > 0$.

	A	B	C
	1.961	1.961	1.961
β_1	>	>	>
β_2	>	>	>
β_3	>	>	>
β_4	>	>	>
β_5	>	>	>
β_6		>	>
β_7		>	>
β_8		>	>
β_9		>	>
β_{10}			>
β_{11}			>
β_{12}			>

Tabell 6: Resultat fra t-test av resultatene for Norge, ved et 5% signifikansnivå.

Som vi ser er $|TS| >$ kritisk verdi for alle estimatorene, noe som betyr at vi med 95% statistisk sannsynlighet kan si at estimatorene er ulik null. Det betyr at variablene er av betydning i sine respektive modeller. Dette betyr at modell B og C antyder at gutter har mindre effekt av mange barnebøker i hjemmet og foreldrenes yrke enn jenter, mens jenter har mindre effekt av mange bøker i hjemmet og foreldrenes utdanning. Vi kan se nærmere på dette ved å utføre en F-test, hvor vi ser på to variabler samtidig. Som beskrevet i 6.4 baserer F-testen seg på endringer i residualkvadratsummen ved å legge på restriksjoner. Restriksjonene vi legger til her er om både estimatoren foran *gutt* og et av interaksjonsleddene er lik null. I testobservatoren bruker vi determinasjonskoeffisienten R^2 fra regresjon med og uten disse restriksjonene. R^2 med restriksjoner får vi fra Tabell 5, mens R^2 med restriksjoner får vi fra regresjonsanalyse uten variablene. Sistnevnte er presentert i Appendiks A. Vi benytter oss av signifikansnivå $\alpha = 5\%$. Vi bruker Stata til å finne den kritiske verdien av f-distribusjonen ved å bruke kommandoen `di invFtail`, med frihetsgrad tilsvarende antall behandlingsgrupper (2) og utvalgsstørrelse N minus antall variabler K. Kritisk verdi, i kursiv, og resultatene fra F-testen er presentert i Tabell 7. Her representerer $>$ at absoluttverdien av TS er større enn kritisk verdi for den korresponderende estimatorene, som betyr at vi kan forkaste nullhypotesen til fordel for den alternative hypotesen. $<$ representerer at absoluttverdien av TS er mindre enn kritisk verdi, som betyr at vi ikke kan forkaste nullhypotese til fordel for den alternative hypotesen. Estimatorene tilsvarer slik de presenteres i A, B, og C. I alle tilfellene er nullhypotesen at begge estimatorene er null, og den alternative hypotesen er at $\beta_i, \beta_j > 0$.

	β_5 fra B	β_5 fra C
	2.999	2.999
β_6	>	<
β_7	>	<
β_8	>	<
β_9	>	<

Tabell 7: Resultat fra F-test av resultatene for Norge, ved et 5% signifikansnivå.

For modell B kan vi altså forkaste nullhypotesen til fordel for den alternative hypotesen. F-testen forteller oss at minst én av variablene skal være ulik null med 95% statistisk sannsynlighet, og at påvirkningen fra begge variablene er statistisk signifikant. Vel å merke kan vi ikke konkludere noe om hver enkelt av variablene. Når det gjelder C, kan vi imidlertid ikke forkaste nullhypotesen. Dette innebærer at F-testen ikke viser en statistisk signifikant sammenheng mellom variablene, og at hverken en eller begge variablene har en statistisk signifikant påvirkning på responsvariabelen ved en konfidensnivå på 95

8.2 Resultater fra regresjonsanalyse for Brasil

Resultatene fra regresjonsanalysen for Brasil er presentert i Tabell 8. Resultatene viser at variablene *profesjonell_jobb*, *høyere_utdanning*, *mange_bøker*, og *mange_barnebøker* alle har positive og signifikante ($p < 0.05$) estimater. Dette indikerer at barn med foreldre med profesjonell jobb og høyere utdanning, samt med tilgang på mange bøker og barnebøker i hjemmet har bedre leseferdigheter. Dette indikerer at det vi omtaler som sosial klasse” har en positiv effekt på leseferdighetene. Resultatene viser også at elever som ikke er født i Norge og elever med mye fravær, har dårligere leseferdigheter - dette er som forventet. Effekten av å prate testens språk i hjemmet har som forventet en positiv effekt på elevens leseferdigheter, men effekten er ikke statistisk signifikant.

I resultatene for Brasil er variabelen for kjønn negativ og signifikant ($p < 0.05$), noe som indikerer at mannlige elever har lavere leseferdigheter sammenlignet med kvinnelige elever.

Når vi ser på interaksjonstermene som inkluderer *gutt* variabelen ser vi at *profesjonell_jobb_gutt* og *mange_bøker_gutt* ikke er signifikante. Derimot er *profesjonell_jobb_gutt* positiv og signifikant ($p < 0.1$), som tilsier at effekten av foreldrenes yrke er større hos gutter enn hos jenter i Brasil. Også *mange_barnebøker_gutt* er signifikant ($p < 0.05$), og negativ. Dette antyder at mannlige elever opplever en mindre positiv effekt av å ha mange barnebøker sammenlignet med kvinnelige elever.

Introduksjonen av interaksjonstermene i modell B førte til reduksjon i effekten av *høyere_utdanning* og *mange_bøker*, mens effekten av *mange_barnebøker* og *gutt* økte. Effekten av *profesjonell_jobb* holdt seg relativt lik. Introduksjonen av kontrollvariablene i modell C dempet disse effektene, med unntak av for *høyere_utdanning*, som ble ytteligere redusert.

R^2 er relativt lav, men indikerer likevel at modellene forklarer en betydelig del av variasjonen i leseferdighetene til elevene i Brasil.

Variabler	A	B	C
	<i>lese</i>	<i>lese</i>	<i>lese</i>
<i>profesjonell_jobb</i>	33.36828* (6.54)	35.88438* (4.83)	33.97687* (4.64)
<i>høyere_utdanning</i>	67.77273* (12.35)	56.32134* (7.07)	53.82199* (6.85)
<i>mange_bøker</i>	15.23759* (2.46)	6.536498 (0.76)	4.127513 (0.46)
<i>mange_barnebøker</i>	32.09991* (4.93)	46.60001* (5.16)	44.66519* (5.02)
<i>gutt</i>	-18.70639* (-4.44)	-23.54537* (-4.07)	-20.08156* (-3.51)
<i>profesjonell_jobb_gutt</i>		-3.362234 (-0.33)	-4.176139 (-0.41)
<i>høyere_utdanning_gutt</i>		21.44465 (1.95)	20.11987 (1.86)
<i>mange_bøker_gutt</i>		18.00103 (1.45)	17.2851 (1.42)
<i>mange_barnebøker_gutt</i>		-29.89814* (-2.30)	-29.77695* (-2.32)
<i>ikke_født</i>			-46.32653* (-2.20)
<i>mye_fravær</i>			-40.34801* (-7.97)
<i>språk_hjemme</i>			22.2485 (0.78)
konstant	420.8831* (116.95)	423.1482* (103.08)	412.0619* (14.39)
<i>N</i>	2,167	2,167	2,167
<i>R</i> ²	0.2393	0.2425	0.2657

Tabell 8: Resultat fra regresjonsanalyse for Brasil.

* $p < 0.05$, t-score i parentes.

For å sjekke om estimatorene foran interessevariablene er av signifikant betydning gjennomfører vi hypotesetester for hver av variablene. Vi benytter oss av teorien beskrevet i kapittel 6.4 og robuste standardavvik. For å teste enkeltparametre bruker vi t-test. Her ønsker vi å se på om estimatorene foran variablene er signifikant ved et 5% signifikansnivå. Her brukes en tosidig test, og derfor bruker vi $\alpha = 5\%/2 = 2.5\%$. Vi bruker Stata til å finne den kritiske verdien av t-distribusjonen ved å bruke kommandoen `invttail`, frihetsgrad tilsvarende utvalgsstørrelse N minus antall variabler K minus 1. Kritisk verdi, i kursiv, og resultatene fra t-testen er presentert i Tabell 9. Her representerer $>$ at absoluttverdien av TS er større enn kritisk verdi for den korresponderende estimatorene, som betyr at vi kan forkaste nullhypotesen til fordel for den alternative hypotesen. $<$ representerer at absoluttverdien av TS er mindre enn kritisk verdi, som betyr at vi ikke kan forkaste nullhypotese til fordel for den alternative hypotesen. Estimatorene tilsvarer slik de presenteres i A, B, og C. I alle tilfellene er nullhypotesen at verdien av estimatoren er lik null, og den alternative hypotesen er at $\beta_i > 0$.

	A	B	C
	1.961	1.961	1.961
β_1	>	>	>
β_2	>	>	>
β_3	>	>	>
β_4	>	>	>
β_5	>	>	>
β_6		>	>
β_7		>	>
β_8		>	>
β_9		>	>
β_{10}			>
β_{11}			>
β_{12}			>

Tabell 9: Resultat fra t-test av resultatene for Brasil, ved et 5% signifikansnivå.

Som vi ser er $|TS| >$ kritisk verdi for alle estimatorene, noe som betyr at vi med 95% statistisk sannsynlighet kan si at estimatorene er ulik null. Det betyr at variablene er av betydning i sine respektive modeller. Dette gjelder altså også for interaksjonsleddene. Dette betyr at modell B og C antyder at gutter har mindre effekt av mange barnebøker i hjemmet og foreldrenes yrke enn jenter, mens jenter har mindre effekt av mange bøker i hjemmet og foreldrenes utdanning. Vi kan se nærmere på dette ved å utføre en F-test, hvor vi ser på to variabler samtidig. Som beskrevet i 6.4 baserer F-testen seg på endringer i residualkvadratsummen ved å legge på restriksjoner. Restriksjonene vi legger til her er om både estimatoren foran *gutt* og et av interaksjonsleddene er lik null. I testobservatoren bruker vi determinasjonskoeffisienten R^2 fra regresjon med og uten disse restriksjonene. R^2 med restriksjoner får vi fra Tabell 8, mens R^2 med restriksjoner får vi fra regresjonsanalyse uten variablene. Sistnevnte er presentert i Appendiks A. Vi benytter oss av signifikansnivå $\alpha = 5\%$. Vi bruker Stata til å finne den kritiske verdien av f-distribusjonen ved å bruke kommandoen `di invFtail`, med frihetsgrad tilsvarende antall behandlingsgrupper (2) og utvalgsstørrelse N minus antall variabler K. Kritisk verdi, i kursiv, og resultatene fra F-testen er presentert i Tabell 10. Her representerer $>$ at absoluttverdien av TS er større enn kritisk verdi for den korresponderende estimatorene, som betyr at vi kan forkaste nullhypotesen til fordel for den alternative hypotesen. $<$ representerer at absoluttverdien av TS er mindre enn kritisk verdi, som betyr at vi ikke kan forkaste nullhypotesen til fordel for den alternative hypotesen. Estimatorene tilsvarer slik de presenteres i A, B, og C. I alle tilfellene er nullhypotesen at begge estimatorene er null, og den alternative hypotesen er at $\beta_i, \beta_j > 0$.

	β_5 fra B	β_5 fra C
	2.999	2.999
β_6	>	>
β_7	>	>
β_8	>	>
β_9	>	>

Tabell 10: Resultat fra F-test av resultatene for Brasil, ved et 5% signifikansnivå.

For både modell B og C kan vi altså forkaste nullhypotesen til fordel for den alternative hypotesen. F-testen forteller oss at minst én av variablene skal være ulik null med 95% statistisk sannsynlighet, og at påvirkningen fra begge variablene er statistisk signifikant. Vel å merke kan vi ikke konkludere noe om hver enkelt av variablene.

8.3 Sammenligning

Dersom vi sammenligner resultatene i Tabell 5 og 8 ser vi at vår analyse avdekker store likheter og forskjeller ved Norge og Brasil i forbindelse med variabler som påvirker leseferdigheter på PIRLS-undersøkelsen. Vi observerer at sosial klasse, definert av foreldrenes arbeidsforhold og utdanningsnivå, samt bøker i hjemmet, påvirker elevprestasjonene i begge land. I hvor stor grad varierer dog mellom de landene, og resultatene tilsier at sosioøkonomisk bakgrunn har en større betydning i Brasil enn i Norge.

Variabelen *profesjonell_jobb* har en positiv effekt på leseferdigheter i begge land, men effekten er tilnærmet dobbelt så stor i Brasil som i Norge. Høyere utdanning hos foreldrene gir en positiv effekt i begge landene, og effekten er tre ganger så stor i Brasil som i Norge. Både antall bøker og barnebøker i hjemmet har en positiv effekt i begge land, men effekten av antall ”vanlige” bøker i hjemmet har en høyere effekt i Norge, mens antall barnebøker har mer enn dobbelt så stor effekt i Brasil sammenlignet med Norge. Gutter presterer dårligere enn jenter i begge land, men effekten er større i Brasil. Interaksjonsleddene for kjønn og profesjonell jobb, og kjønn og høyere utdanning viser de samme tendensene i begge land, dvs. gutter opplever en mindre effekt av profesjonell jobb, og større effekt av høyere utdanning, enn jenter. Effekten av *profesjonell_jobb_gutt* har mindre effekt i Brasil, mens effekten av *høyere_uttanning_gutt* er høyere i Brasil, sammenlignet med Norge. I Norge opplever gutter en mindre effekt av antall bøker i hjemmet sammenlignet med jenter, mens de opplever en større effekt av antall barnebøker i hjemmet. I Brasil er det motsatt. Effekten av å ikke være født i landet og å ha mye fravær er signifikant og negativ i begge land, men betydelig større i Brasil. Elever som snakker testens språk i hjemmet opplever en positiv effekt av dette i Brasil, men en negativ effekt i Norge.

For begge land er R^2 er relativt lav, men betraktelig lavere for resultatene for Norge enn for Brasil. Likevel er det grunnlag for å påstå at modellene forklarer en betydelig del av variasjonen i leseferdighetene til elevene. Vi kan merke oss at Modell C har høyeste R^2 for begge land. Dette kan indikere at denne modellen forklarer en større del av variasjonen. Dette antyder at inkluderingen av de tre kontrollvariablene bidrar til en bedre modell. Det er imidlertid viktig å være oppmerksom på at selv om R^2 kan øke når flere variabler legges til, betyr ikke nødvendigvis dette at kontrollvariablene er relevante. Det er mulig at inkluderingen av flere variabler kan føre til overjustering eller en modell som er for kompleks i forhold til den underliggende virkeligheten.

9 Diskusjon

Dersom vi sammenligner våre resultater med tidligere litteratur presentert i kapittel 4.1 finner vi at resultatene fra vårt arbeid underbygger tidligere litteratur på feltet. For eksempel peker våre resultater på en positiv effekt av foreldres utdanningsnivå og tilgang på bøker i hjemmet på elevprestasjoner, i likhet med i Colemans rapport [2]. Også i rapportene ”Why Poor Children Are More Likely to Become Poor Readers: The School Years” og ”Monitoring Australian Year 4 student achievement internationally: TIMSS and PIRLS 2011” pekes det på en tilsvarende positiv effekt av tilganger på bøker i hjemmet på elevprestasjoner som vi har påvist i vårt arbeid [13], [14]. I den Tjekkiske studien The Effect of Parents’ Education and Income on Children’s School Performance: the Mediating Role of the Family Environment and Children’s Characteristics, and Gender Differences pekes det på en direkte og indirekte positiv effekt av foreldrenes utdanning på barnas prestasjoner, som vi også finner igjen i vårt arbeid [12].

Den relativt store effekten av foreldrenes yrke og utdanning i Brasil underbygger vår initiale teori om at de positive effekten av høy sosial klasse er større i Brasil, hvor de sosiale forskjellene er større. Forskjellen på effekten av tilganger på ulike typer bøker i hjemmet kan muligens forklares med kulturelle forskjeller og ulikheter i lesepraksis mellom landene.

Gjennomgående ser vi også at kjønnsforskjeller spiller en rolle, med gutter som generelt presterer dårligere enn jenter i begge landene. Imidlertid ser vi at kjønnsforskjellene er mer uttalt i Brasil enn i Norge, noe som indikerer at kjønnsforskjellene i elevprestasjoner kan være mer fremtredende i land med større sosiale forskjeller.

9.1 Videre arbeid

Mulige utvidelser for vår analyse kan være å utvide med en ikke-lineær modell. Dette kan gjøres ved å introdusere en variabel som er den kvadrerte av en av de eksisterende variablene for sosial klasse. Dette ville gi oss muligheten til å undersøke eventuelle ikke-lineære sammenhenger mellom den eksisterende variabelen og responsvariabelen. Dette kan være spesielt nyttig hvis det er indikasjoner på at effekten av sosial klasse på responsvariabelen ikke er lineær. For å tydeligere hvordan effekten av sosial klasse varierer mellom ulike land med ulik grad av sosial ulikhet kan en også inkludere flere land i analysen.

10 Oppsummering og konklusjon

I oppgaven vår har vi forsøkt å analyse hvorvidt det er mulig å observere betydningen av sosial klasse på elevprestasjoner i Brasil og Norge, og studere hvordan disse effektene varierer mellom to land med henholdsvis store og små sosiale forskjeller. I mangel på tilgang på informasjon om familiens inntekt, har sosial klasse her blitt definert ut ifra foreldrenes yrke og utdanning, samt tilgangen på bøker og barnebøker i elevens hjem. I vår analyse har vi brukt datamateriale hentet fra utvalgsundersøkelsen Progress in International Reading Literacy Study, PIRLS. Modellene våre er utledet med utgangspunkt i skoleproduktfunksjonen. I tillegg til å se på effekten av sosial klasse, har vi undersøkt effekten kjønn har på elevens prestasjon. Vi introduserte også kontrollvariabler, for å undersøke hvorvidt eleven er født i landet testen utføres i, om eleven har mye fravær, og språket eleven snakker i hjemmet påvirker resultatene vi er ute etter.

I analysen har vi brukt minste kvadraters metode, og teorien som er benyttet er presentert. Vi begynte med en enkel modell som bestod av vår avhengige variabel, *lese*, og våre fire interessevariabler, *høyere_uttanning*, *profesjonell_jobb*, *mange_bøker*, og *mange_barnebøker*. Deretter utvidet vi modellen ved å introdusere interaksjonsledd, mellom *gutt* og de fire interessevariablene. Til slutt utvidet vi med kontrollvariablene.

Vi har tolket resultatene fra analysen, og derav konkludert med at i våre modeller observerer vi stor påvirkning av sosial klasse på elevens prestasjoner. Vi observerer også at denne påvirkningen er større i Brasil enn i Norge. Resultatene underbygger det tidligere arbeid har konkludert med i lignende analyser. Dog evner ikke vår analyse å fange effekten av sosial klasse fullstendig, ettersom variabelen for inntekt ikke var tilgjengelig i datasettet.

Videre arbeid burde undersøke ikke-lineære sammenhenger mellom sosial klasse og elevprestasjoner, samt inkludere flere land for å bedre reflektere hvordan effektene varierer avhengig av sosiale forskjeller i landet.

A R^2 for bruk i F-test

Her presenteres determinasjonskoeffisienten R^2 fra regresjon utført uten variabelen *gutt*, og enten *profesjonelljobb_gutt*, *høyere_utdanning_gutt*, *mange_bøker_gutt*, eller *mange_barnebøker_gutt*. Tabellen viser R^2 for modell B og C, ettersom det kun er disse modellene som inkluderer interaksjonsleddene. Tabell 11 og 12 viser henholdsvis disse verdiene for Norge og Brasil.

Ekskluderte variabler	R^2 for B	R^2 for C
<i>gutt</i> og <i>profesjonelljobb_gutt</i>	0.0745	0.0915
<i>gutt</i> og <i>høyere_utdanning_gutt</i>	0.0751	0.0922
<i>gutt</i> og <i>mange_bøker_gutt</i>	0.0748	0.0922
<i>gutt</i> og <i>mange_barnebøker_gutt</i>	0.0752	0.0926

Tabell 11: Determinasjonskoeffisienten R^2 fra regresjon utført uten restriksjoner for modellene B og C, til bruk i F-test for Norge.

Ekskluderte variabler	R^2 for B	R^2 for C
<i>gutt</i> og <i>profesjonelljobb_gutt</i>	0.2357	0.2605
<i>gutt</i> og <i>høyere_utdanning_gutt</i>	0.2369	0.2611
<i>gutt</i> og <i>mange_bøker_gutt</i>	0.2370	0.2611
<i>gutt</i> og <i>mange_barnebøker_gutt</i>	0.2354	0.2594

Tabell 12: Determinasjonskoeffisienten R^2 fra regresjon utført uten restriksjoner for modellene B og C, til bruk i F-test for Brasil.

Bibliografi

- [1] E. Løkeland-Stai. «Fillerister høyere utdanning i USA: “Det er bedre å bli født rik enn å bli født smart.”» (2020), adresse: <https://www.khrono.no/fillerister-hoyere-utdanning-i-usa-det-er-bedre-a-bli-fodt-rik-enn-a-bli-fodt-smart/493608> (sjekket 5. feb. 2024).
- [2] J. S. Coleman, *Equality of Educational Opportunity*. Washington: U.S. Department of Health, Education, og Welfare, 1966.
- [3] C. W. Beck, «Sosial Bakgrunn - et Skoleproblem», *Bedre Skole*, nr. 2, apr. 2012. adresse: <https://utdanningsforskning.no/artikler/2012/sosial-bakgrunn---et-skoleproblem/>.
- [4] F. India. «Top 10 Largest Economies in the World in 2024». Hentet 18. Apr. 2024. (2024), adresse: <https://www.forbesindia.com/article/explainers/top-10-largest-economies-in-the-world/86159/1>.
- [5] The World Bank, *Open Data DataBank*, 2022. adresse: <https://data.worldbank.org/>.
- [6] J. Eide og K. Stæhr, «Brasils Finanskrisen I et Historisk Perspektiv - Statsfinanser, Inflasjon Og Stabilisering», s. 217–229, 1999, Hentet 18. Apr. 2024. adresse: <https://norges-bank.brage.unit.no/norges-bank-xmlui/handle/11250/2480399>.
- [7] K. Martinsen, «Renteendringer Har En Fordelingseffekt», *Finansavisen*, jul. 2021, Hentet 18. Apr. 2024. adresse: https://www.finansavisen.no/nyheter/debattinnlegg/2021/07/16/7706388/renteendringer-har-en-fordelingseffekt?zephrr_sso_ott=sOIUTe.
- [8] R. Reeves og S. Nzau. «Poverty Hurts the Boys the Most: Inequality at the Intersection of Class and Gender», Brookings Institution. (2021), adresse: <https://policycommons.net/artifacts/4142992/poverty-hurts-the-boys-the-most/4950869/>.
- [9] L.-E. Borge og M. Rønning, «Bidrar økt ressursbruk i grunnskolen til bedre elevprestasjoner?», *Statistiske analyser*, årg. 2009, s. 101–112, 2009, ISSN: 0804-3221.
- [10] D. Ingebø, E. S. Foss og N. Fazli, «Innvandrere – hvordan går det med dem i utdanningsløpet?», Statistisk sentralbyrå, tekn. rapp. 2021/37, 2021.
- [11] S. Stumm, S. Cave og P. Wakeling, «Persistent association between family socioeconomic status and primary school performance in Britain over 95 years», *npj Science of Learning*, årg. 7, apr. 2022. DOI: 10.1038/s41539-022-00120-3.
- [12] P. Raudenská og D. Hamplová, «The Effect of Parents’ Education and Income on Children’s School Performance: the Mediating Role of the Family Environment and Children’s Characteristics, and Gender Differences», *Polish Sociological Review*, årg. 218, nr. 2, s. 247–272, 2022. DOI: 10.26412/psr218.06.
- [13] J. Buckingham, K. Wheldall og R. Beaman-Wheldall, «Why Poor Children Are More Likely to Become Poor Readers: The School Years», *Australian Journal of Education*, årg. 57, nr. 3, s. 190–213, 2013. DOI: 10.1177/0004944113495500. adresse: <https://doi.org/10.1177/0004944113495500>.
- [14] S. Thomson, K. Hillman, N. Wernert, M. Schmid, S. Buckley og A. Munene, *Monitoring Australian Year 4 student achievement internationally: TIMSS and PIRLS 2011*. jan. 2012.
- [15] J.-C. Croizet og M. Millet, «Social Class and Test Performance From Stereotype Threat to Symbolic Violence and Vice Versa», i jan. 2012, s. 188–201, ISBN: 9780199732449. DOI: 10.1093/acprof:oso/9780199732449.003.0012.
- [16] M. Carnoy, L. Rosa og A. Simões, «Trends in the academic achievement gap between high and low social class children: The case of Brazil», *International Journal of Educational Development*, årg. 94, s. 102650, 2022, ISSN: 0738-0593. DOI: <https://doi.org/10.1016/j.ijedudev.2022.102650>. adresse: <https://www.sciencedirect.com/science/article/pii/S0738059322001006>.
- [17] H. Bonesrønning, «Utforming av utdanningspolitikken – Hva kan økonomene bidra med?», *Økonomisk Forum*, årg. 58, nr. 3, s. 14–23, 2004.
- [18] OECD, *Education in Brazil: An International Perspective*. OECD Publishing, 2021.
- [19] J. Epland og N. Tuv, *Slik Måler SSB Ulikhet*, SSB, mai 2019. adresse: <https://www.ssb.no/inntekt-og-forbruk/inntekt-og-formue/statistikk/inntekts-og-formuesstatistikk-for-husholdninger/artikler/slik-maler-ssb-ulikhet>.

-
- [20] FN-Sambandet. «Norge». (2021), adresse: <https://fn.no/land/norge> (sjekket 1. mai 2024).
- [21] European Commission. «Organisation of the Education System and of Its Structure». (2023), adresse: <https://eurydice.eacea.ec.europa.eu/national-education-systems/norway/organisation-education-system-and-its-structure> (sjekket 1. mai 2024).
- [22] Encyclopedia Britannica. «Brazil - Primary and Secondary School». (n.d.), adresse: <https://www.britannica.com/place/Brazil/Primary-and-secondary-school> (sjekket 1. mai 2024).
- [23] A. Roman, «A Closer Look Into Brazil's Healthcare System: What Can We Learn?», *Cureus*, årg. 15, nr. 5, e38390, mai 2023. DOI: 10.7759/cureus.38390.
- [24] FN-Sambandet. «Brasil». (2024), adresse: <https://fn.no/land/brasil> (sjekket 1. mai 2024).
- [25] C. N. Baker, «Kulturelle Verdier i Kommunikasjonen», *NDLA*, jun. 2022. adresse: <https://ndla.no/article/35075>.
- [26] W. Hansen og M. Andre, «Hofstedes Kulturelle Dimensjoner Og Pragmatisk Nyttverdi: Svakheter Og Styrker Ved Hans Teoretiske Modell et Litteraturstudie Om Hofstedes Modell Og Relaterte Problemstillinger Knyttet Til Multinasjonalt Management», *Munin.uit.no*, nov. 2020, Hentet 18. Apr. 2024. adresse: <https://munin.uit.no/handle/10037/20799>.
- [27] The Culture Factor, *Country comparison tool*, 2023. adresse: <https://www.hofstede-insights.com/country-comparison-tool?countries=brazil%2Cnorway>.
- [28] G. M. Faitar, «Individualism versus Collectivism in Schools», *College Quarterly*, årg. 9, nr. 4, 2004. adresse: <https://collegequarterly.ca/2006-vol09-num04-fall/faitar.html> (sjekket 24. apr. 2024).
- [29] K. Cortina, S. Arel og J. Smith-Darden, «School Belonging in Different Cultures: The Effects of Individualism and Power Distance», *Frontiers in Education*, årg. 2, s. 56, nov. 2017. DOI: 10.3389/educ.2017.00056.
- [30] D. Figlio, P. Giuliano, U. Özek og P. Sapienza, «Long-Term Orientation and Educational Performance», *American Economic Journal: Economic Policy*, årg. 11, nr. 4, s. 272–309, nov. 2019. DOI: 10.1257/pol.20180374. adresse: <https://doi.org/10.1257/pol.20180374>.
- [31] Utdanningsdirektoratet. «Den internasjonale undersøkelsen PIRLS». [Tilgjengelig online; sist besøkt 4. mars 2024]. (2023), adresse: <https://www.udir.no/tall-og-forskning/internasjonale-studier/pirls/>.
- [32] D. Caldeirinha, *Brazil Performs Poorly in International Reading Test, Focusing on Literacy Is the Way out - News - Lemann Foundation*, Fundação Lemann, okt. 2023. adresse: <https://fundacaolemann.org.br/en/news/brazil-performs-poorly-in-international-reading-test-focusing-on-literacy-is-the-way-out>.
- [33] I. A. for the Evaluation of Educational Achievement, «Progress in International Reading Literacy Study 2021 Report», 2021. adresse: <https://www.iea.nl/data-tools/repository/pirls>.
- [34] R. L. Thomas, *Using Statistics in Economics*, 1. utg. McGraw Hill Higher Education, 2005.
- [35] J. M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7. utg. Cengage Learning, 2019.
- [36] E. Lewinson, *A Comprehensive Guide to Interaction Terms in Linear Regression*, NVIDIA Technical Blog, apr. 2023. adresse: <https://developer.nvidia.com/blog/a-comprehensive-guide-to-interaction-terms-in-linear-regression/>.

