

En empirisk analyse av effekten geografisk bosted har på deltagelse i høyere utdanning

Silje Rørtveit Mundal



**Masteroppgave i samfunnsøkonomi
Februar 2007**

**Institutt for økonomi
Universitet i Bergen**

Forord

Ferdig! Fantastisk å sette punktum for en periode preget av kronisk dårlig samvittighet og mye hardt arbeid. Til tross for dette har arbeidet med masteroppgaven også gitt mye glede. Få ting kan sammenlignes med å få STATA til å danse etter sin pipe, eller lykken over å se nye sammenhenger.

Jeg vil gjerne rette en stor takk for all hjelp til min veileder Arild Aakvik.

Gunhild Tinnmannsvik, Eirin Eikefjord og Inge Mundal fortjener også takk for korrekturlesing, god hjelp og viktig støtte på slutten.

I tillegg til de som konkret har hjulpet meg med oppgaven, vil jeg rette en generell takk til ansatte og medstudenter på institutt for økonomi for en veldig hyggelig studietid. Venner og familie fortjener også ros for å ha holdt ut det evige gnålet om denne oppgaven.

Bergen, februar 2007

Silje Rørtveit Mundal

Sammendrag

Denne masteroppgaven søker å belyse hvilken effekt geografisk avstand mellom hjemsted og nærmeste høyere utdanningsinstitusjon har på individenes deltagelse i høyere utdanning. For å besvare problemstillingen benytter jeg registerdata fra Statistisk Sentralbyrå om kohorten født i 1968 og foreldrene deres. I databehandlingen bruker jeg statistikkverktøyet STATA. På grunn av problemer med multikollinearitet lykkes det ikke å besvare hovedproblemstillingen. De øvrige resultatene fra oppgaven er tradisjonelle, og viser en positiv samvariasjon mellom foreldrenes utdanning, inntekt og alder og sannsynligheten for at barnet har høyere utdanning. Antall søsken påvirker denne sannsynligheten negativt. Det er grunn til å anta at samvariasjonen mellom barnets sannsynlighet for høyere utdanning og de viktigste forklaringsvariablene som er foreldrenes utdanning og inntekt, ikke er kausale.

KAPITTEL 1: INNLEDNING	1
1.1 BAKGRUNN	1
1.2 PROBLEMSTILLING	1
1.3 VIKTIGSTE LITTERATUR	2
1.4 DATA OG METODE	2
1.5 HOVEDRESULTAT	2
1.6 OPPBYGGING AV OPPGAVEN	2
KAPITTEL 2: TIDLIGERE LITTERATUR	4
2.1 UTDYPING AV EVNEBEGREPET	4
2.2 FORKLARINGSVARIABLENE	5
2.2.1 Foreldrenes utdanning	5
2.2.2 Inntekt	7
2.2.3 Familie.....	9
2.2.4 Nabolag og skole	10
2.2.5 Avstand fra hjemsted til høyere utdanningsinstitusjon	11
2.3 SVAKHETER VED TIDLIGERE FORSKNING	12
2.3.1 Modelladopsjon uten teoretisk fundament	13
2.3.2 Måleproblemer, stedfortredende variabler og instrumentvariabler	13
Instrumentvariabler	14
2.3.3 Manglende sammenlignbarhet.....	15
2.3.4 Forveksling av samvariasjon og kausalitet.....	16
KAPITTEL 3: DATABESKRIVELSE	17
3.1 DATASETT	17
3.1.1 Utdanning.....	18
3.1.2 Inntekt.....	20
3.1.3 Familiekarakteristika.....	22
3.1.4 Avstand mellom hjemsted og nærmeste høyere utdanningsinstitusjon	22
3.2 DESKRIPTIV STATISTIKK.....	24
3.2.1 Sammenheng mellom familiekarakteristika og barnas utdanning.....	25
KAPITTEL 4: EMPIRISKE LIGNINGER	35
4.1 REGRESJONSMODELLER	35
4.1.1 LINEÆRE SANNSYNLIGHETSMODELLER	35
4.2 TESTER AV REGRESJONENE	37
4.2.1 Multikollinearitet.....	37
4.2.2 Chow-test.....	38
4.2.3 Heteroskedastisitet.....	39
4.3 ESTIMERTE MODELLER	40
KAPITTEL 5: RESULTATER	42
5.1 REGRESJONSRESULTATER	42
5.1.1 Modell 1.....	42
5.1.2 Modell 2.....	45
5.1.3 Modell 3.....	47
5.2 SELEKSJON VERSUS KAUSALITET	50
5.3 GRAD AV SOSIAL MOBILITET	51
KAPITTEL 6: AVSLUTNING	54
6.1 KONKLUSJON	54
6.2 KRITIKK OG VIDERE FORSKNING	55
LITTERATURLISTE	57
APPENDIKS	59
APPENDIKS A, ØKONOMETRI	59
APPENDIKS B, REGRESJONSRESULTATER	60
APPENDIKS C, GRAFER	61

Tabelloversikt

Tabell 3.1: Kvalitativt innhold av utdanningsnivå	s. 12
Tabell 3.2: Oversikt over inntektsvariablene som benyttes	s. 21
Tabell 3.3: Deflaterte og diskonterte familieinntekter, utvalgte år	s. 21
Tabell 3.4: Prosentvis fordeling av antall barn i familiene i 1970	s. 22
Tabell 3.5: Deskriptiv statistikk, menn	s. 24
Tabell 3.6: Deskriptiv statistikk, kvinner	s. 24
Tabell 3.7: Sammenheng mellom familiekarakteristika og høyere utdanning	s. 25
Tabell 4.1: Korrelasjoner mellom variablene i modell 1 og modell 3	s. 37
Tabell 4.2: Korrelasjoner mellom inntektsvariablene	s. 38
Tabell 5.1: Resultater fra modell 1. Lineær estimering.	s. 42
Tabell 5.2: Resultater fra modell 1. Probitestimeringen	s. 42
Tabell 5.3: Resultater fra modell 2. Ulike inntektseffekter	s. 45
Tabell 5.4: Resultater fra provisorisk estimering av inntektsvinduer	s. 45
Tabell 5.5: Resultater fra modell 3. Avstand	s.47
Tabell 5.6 a)- f): Resultater fra modell 3. Variabel for variabel	s.48

Figuroversikt

Figur 3.1: Grafisk fremstilling av høyeste fullførte utdanningsnivåbarn og foreldre	s. 19
Figur 3.2: Sammenheng mellom foreldres utdanning og barnas utdanning, menn	s. 27
Figur 3.3: Sammenheng mellom foreldres utdanning og barnas utdanning, kvinner	s. 28
Figur 3.4: Sammenheng mellom familieinntekt og barnas utdanning, menn	s. 29
Figur 3.5: Sammenheng mellom familieinntekt og barnas utdanning, kvinner	s. 30
Figur 3.6: Sammenheng mellom geografisk bosted og barnas utdanning, menn	s. 31
Figur 3.7: Sammenheng mellom geografisk bosted og barnas utdanning, kvinner	s. 32
Figur 3.8: Kumulativ fordeling av antall år utdanning sett i forhold til geografisk bosted, menn	s. 33
Figur 3.9: Kumulativ fordeling av antall år utdanning sett i forhold til geografisk bosted, kvinner	s. 34

KAPITTEL 1: INNLEDNING

1.1 Bakgrunn

Utdanning har en sentral posisjon i det norske samfunnet fordi et visst kunnskapsnivå i befolkningen er en grunnleggende forutsetning for å opprettholde og videreutvikle viktige samfunnsstrukturer. Dette kan konkretiseres med eksempler både på samfunnsmessig og individuelt nivå.

Samfunnsmessig nivå

- De demokratiske institusjonene er bygget opp rundt ideen om enkeltindividets rett til å velge. Skal man fatte valg som er i ens egen interesse, krever dette gode kunnskaper.
- Den norske kulturen føres videre og utvikles ved at hvert enkelt individ får en systematisk opplæring i skoleverket.¹
- I økonomiske termer betegnes mennesket som humankapital. Investering i kunnskap hos befolkningen er følgelig det samme som å bygge opp denne kapitalen, noe som igjen gir økonomiske gevinster i form av mer effektiv produksjon og økonomisk vekst.²

Individnivå

- Utdanning er en nøkkel til sosioøkonomisk suksess for enkeltindividet. I tillegg til å være en av parametrene suksessen måles etter, lykkes individer med høy utdanning ofte også etter andre målestokker som for eksempel høy inntekt og god helse.
- På den annen side er individer med lav utdanning i større grad enn andre assosiert med kriminalitet, arbeidsledighet og dårlig helse. Dette er atferd som er forbundet med høye kostnader for samfunnet.³

Når man studerer hvilke effekter utdanning har på samfunnet, skjønner man at det er fordelaktig med et utdanningsvesen som favner om flest mulig. Dette ønsket har materialisert seg i følgende politiske målsettinger:

¹ For en mer utdypende diskusjon av de to første punktene se Hernes (1976).

² Jamfør Romer (2006).

³ Heckman og Masterov (2004) peker på noen negative effekter ved å ha lav utdanning.

1. Det norske utdanningssystemet skal være blant de beste i verden når det gjelder faglig nivå og bredde i rekrutteringen.
2. Barn og unge skal ha lik rett til utdanning uavhengig av bosted, kjønn, sosial og kulturell bakgrunn.⁴

Målene er et tegn på at det er en tydelig vilje til å utjevne forskjeller i befolkningen, og at det er et politisk ønske om høy sosial mobilitet i forhold til utdanning. Sosial mobilitet er et mål på i hvilken grad sosioøkonomisk suksess avhenger av familiebakgrunn. Fra politisk hold er det altså et ønske om at barnas utdanningsforløp i minst mulig grad skal avhenge av foreldrenes utdanning, inntekt og bosted.

Skal man utvikle gode redskaper for å nå politiske mål krever dette kunnskap om emnet politikken skal implementeres på. For å effektivt kunne nå politiske målsettinger i forhold til utdanning krever dette god kunnskap om hvilke mekanismer som styrer individenes utdanningsforløp. Formålet med denne studien er å øke kunnskapen om hvilke faktorer som påvirker deltagelsen i høyere utdanning for kohorten født i 1968.

1.2 Problemstilling

Mitt konkrete bidrag til å øke kunnskapen på feltet består i å kartlegge effekten geografisk avstand⁵ mellom hjemsted og nærmeste høyere utdanningsinstitusjon har på deltagelsen i høyere utdanning. Denne tilnærmingen er interessant på grunn av det politiske målet om at geografisk bosted ikke skal ha konsekvenser for utdanningsforløpet.

Problemformuleringen jeg har arbeidet etter er derfor:

Hvilken effekt har avstand mellom hjemsted og nærmeste høyere utdanningsinstitusjon på sannsynligheten for at individet tar høyere utdanning?

⁴ Begge disse målene er hentet fra nettsiden til kunnskapsdepartementet 15.02.06

⁵ For å øke leservennligheten vil det noen ganger kun stå avstand, når jeg mener avstand mellom hjemsted og nærmeste høyere utdanningsinstitusjon.

1.3 Viktigste litteratur

Det har vært gjort få studier på sosial mobilitet i forhold til utdanning i Norge, og litteraturen det empiriske arbeidet er forankret i, stammer i stor grad fra utlandet. USA er et foregangsland på denne type forskning, og har satt standarden for hvilke variabler som er vanlig å ha med i regresjonene. Disse bestemmes ut i fra amerikanske forhold, som ikke nødvendigvis er sammenlignbare med norske, blant annet fordi amerikanerne finansierer store deler av høyere utdanning fra egen lomme. På den annen side kan man anta at mange av konseptene i litteraturen er allmenngyldige, og at likhetene mellom rike vestlige land er mange. Artikkene gir derfor viktig input når man skal analysere norske forhold. Heckman, Card og Chevalier er de viktigste utenlandske forfatterne, mens Salvanes, Raaum og Aamodt har viktige bidrag om norske forhold.

1.4 Data og metode

For å besvare problemstillingen benytter jeg regresjonsanalyse som er en metode for å estimere sammenhengen mellom ulike variabler. Datamaterialet jeg bruker er registerdata fra Statistisk sentralbyrå (SSB). Dette inneholder informasjon om variabler som inntekt, utdanning og bosted for kohorten født i 1968 og foreldrene deres. For å fange opp effekten avstand mellom hjemsted og nærmeste høyere utdanningsinstitusjon har på deltagelsen i høyere utdanning, baserer mine analyser seg av praktiske grunner på individer med fylkestilhørighet i Hordaland og Sør-Trøndelag. Statistikkverktøyet som benyttes er STATA.

1.5 Hovedresultat

På grunn av problemer med multikollinearitet lyktes det ikke å estimere effekten avstand har på sannsynligheten for deltagelse i høyere utdanning for denne kohorten. De øvrige resultatene er tradisjonelle. Foreldrenes inntekt og utdanning, er variablene som i størst grad kan benyttes til å forklare forskjeller i sannsynligheten for deltagelse i høyere utdanning for individene. Det er små forskjeller i inntekt mellom familiene. Dette sett i sammenheng med tidligere empiri, tyder på sammenhengene ikke i stor grad kan tilskrives kausalitet.

1.6 Oppbygging av oppgaven

Opgaven består av 6 kapitler. I tillegg til dette innledningskapittelet kan innholdet i oppgaven oppsummeres slik.

Kapittel 2 – Tidligere litteratur

Kapittel 2 er en gjennomgang av relevant litteratur og empiri. Jeg introduserer her forklaringsvariablene det har vært vanlig å benytte i tilsvarende studier, før jeg oppsummerer resultatene. Forskningen er preget av noe uenighet i bruk av metodikk og tolkning av resultater. Disse aspektene er det viktig å ha kunnskap om. Kapitlet avsluttes derfor med en gjennomgang av kritikken som har vært rettet mot tidligere forskning.

Kapittel 3 - Databeskrivelse

Kapittel tre består av utfyllende databeskrivelser og deskriptiv statistikk. For at datamaterialet skal kunne benyttes til å besvare problemstillingene var det nødvendig å foreta noen endringer og tilføyelser. I tillegg til å beskrive dataene, inneholder kapitlet også informasjon om hvilke endringer som er gjort, og begrunnelsen for disse. I den deskriptive delen inngår en mer utdypende databeskrivelse.

Kapittel 4 - Empiriske ligninger

I kapittel fire introduseres regresjonsmodellene som benyttes i det empiriske arbeidet. For å unngå problemer med mulikollinearitet sjekkes korrelasjonene mellom variablene. Ligningene testes for heteroskedastisitet og chow-testen benyttes til å se om det er kjønnsforskjeller som gjør at det må estimeres en modell for kvinner og en for menn.

Kapittel 5 - Resultater

Resultatene presenteres og drøftes i forhold til tidligere forskning. I tillegg evaluerer jeg om sammenhengene mellom variablene skyldes kausalitet eller seleksjon. Jeg anvender også dataene til å diskutere grad av sosial utdanningsmobilitet mellom generasjonene i dette utvalget.

Kapittel 6 - Avslutning

I dette kapitlet presenteres de overordnede konklusjonene. Kapitlet inneholder også kritikk av eget arbeid, samt forslag til videre forskning.

KAPITTEL 2: TIDLIGERE LITTERATUR

Kapittelet oppsummerer relevant litteratur om hvilke faktorer som kan forklare individenes utdanningsforløp. Forskningen er preget av få teoretiske modeller. De empiriske ligningene som benyttes er ofte basert på "common sense", og variablene man har tilgang på. Jeg fokuserer derfor på å formidle begrunnelsene forskningen har benyttet for å rettferdiggjøre inklusjon av variablene i regresjonene.

Evner er et sentralt begrep i denne type forskning, fordi man antar at det er dette som i hovedsak styrer utdanningsforløpet. Kapittelet starter derfor med en utdyping av evnebegrepet. Jeg går deretter gjennom de forklaringsvariablene som har stått sentralt i litteraturen, og oppsummerer viktige resultater fra tidligere studier. Det er noe uenighet blant forskerne på feltet om bruk av metoder og tolkning av resultater. I videre forskning er dette aspekter det er viktig å være oppmerksom på. Kapittelet avsluttes derfor med en gjennomgang av kritikken som har vært rettet mot tidligere studier.

2.1 Utdyping av evnebegrepet

Sosioøkonomisk suksess avhenger av kognitive og ikke-kognitive evner (Heckman, 2005). Med kognitive evner menes hvor intelligent du er, og evnetypen inkluderer blant annet IQ og talent for læring. I følge Heckman (2005) utvikles denne type evner svært tidlig. Allerede i syvårsalderen vil prøver på kognitive evner gi en indikasjon på hvordan individet vil klare seg i arbeidsmarkedet som voksen (Todd og Wolpin, 2003). I fjortenårsalder er intelligens, målt ved hjelp av IQ-tester, ferdig utviklet (Caneiro og Heckman, 2002).

Ikke-kognitive evner beskriver personlighetstrekk som motivasjon, evne til konsentrasjon og innsatsvilje. Dette påvirker utdanning direkte og indirekte. Motiverte individer har større mulighet til å gjennomføre en krevende utdanning enn en tilsvarende personer uten samme motivasjon. Ikke-kognitive evner kan også gi en indirekte effekt siden læring baserer seg på tidligere læring. Siden den motiverte personen har skaffet seg gode forkunnskaper vil han/hun ha lettere for å lære mer. Ikke-kognitive evner er derfor viktig for å fullt ut nytte potensialet i kognitive evner. I følge Heckman (2005) utvikles denne evnetypen i slutten av tenårene.

Evner utvikles i et samspill mellom arv og miljø, men man har ikke eksakt kunnskap om hvordan hver av komponentene påvirker utviklingen. Ulike retninger innenfor fagfeltet har derfor forskjellig innfallsvinkel på betydningen av arv versus miljø. Dette preger det teoretiske fundamentet for modellene. Deler av forskningen baserer seg på arv når evner og utdanning skal forklares, mens andre er mest opptatt av oppvekstmiljø. Det er likevel sterk konsensus om at både arv og miljø har betydning for individets evneutvikling, og det er derfor utfyllende heller, enn problematisk for kunnskapen om emnet, at litteraturen har så ulike innfallsvinkler. At kunnskapen om evner og evneutvikling er så diffus, er likevel et problem for kvantitativ forskning. Siden evner er vanskelig å måle og man er usikker på hvilke faktorer som fremmer evneutvikling, er det vanskelig å finne relevante variabler til regresjonene.

På grunn av dette er variablene som inkluderes i regresjonene med, fordi man antar at dette er faktorer som påvirker ressurstilfanget, som igjen er med på å utvikle evner. Jo større tilgangen er til menneskelige og materielle ressurser, jo større er sannsynligheten for at barnet fullt ut kan nytte sitt medfødte evnepotensial. I denne sammenhengen innebærer ressurser blant annet at foreldre kan/velger å bosette seg i nabolag hvor klimaet er godt for læring, har positive holdninger til utdanning, stimulerer barna sine til utvikling, og investerer i materielle ting som fremmer evneutviklingen. Dette kan for eksempel være bøker og stimulerende leker.

2.2 Forklaringsvariablene

2.2.1 Foreldrenes utdanning

Foreldrenes utdanning kan ha effekt på barnas evner både gjennom arv og miljø. Genetisk overføres evner til barna gjennom mekanismene arv og ektefellevalg. I følge Chevalier et al. (2005) kan medfødt kvalitet på barnet være avgjørende for evneutviklingen. Den medfødte kvaliteten må også gjelde foreldrene, og man kan derfor trekke konklusjon om at høyt utdannede foreldre har fordelaktige gener for læring. Disse arver barna. Foreldrenes utdanning gir på denne måten en indikasjon på barnets genetiske forutsetninger.

Hvem som får barn sammen er ikke tilfeldig. Det er en tendens at høyt utdannede kvinner gifter seg med høyt utdannede menn. Dersom gener er en viktig forutsetning for utvikling av evner, er det naturlig å tenke seg at høyt utdannede par får barn med samme, eller bedre læringsmessige forutsetninger (Behrman og Rosenzweig, 2001).

Oppvekstmiljøet preges av foreldrenes kunnskaper og holdninger. Disse dannes blant annet på bakgrunn av egne erfaringer og utdanning. Foreldre med høy utdanning kan påvirke barnas utdanning ved å tilrettelegge for høyere kvantum av andre innsatsfaktorer som er viktig for barnets evneutvikling (Chevalier et al.,2005).

Sammenlignet med foreldre uten høyere utdanning, kan foreldre med universitets og høyskoleutdanning være mer opptatt av at også barna deres tar høyere utdanning. I den grad det er tilfelle, vil statusverdien av utdanning videreføres fra en generasjon til en annen. Siden det er enklere å lære bort det man kan selv, har foreldre med utdanning større forutsetning for å gi god leksehjelp (Raaum, 2003). Betydningen av mors utdanning har tradisjonelt stått sterkere enn fars, i teori om foreldreutdanning. En kausal forklaring på dette, er at mor oftest er den personen som er mest med barnet i tidlige år når kognitive evner formes.

Resultater fra tidligere forskning

Resultater fra tidligere forskning tyder på at det er en sammenheng mellom barn og foreldres utdanning. Høyt utdannede foreldre får barn som tar høy utdanning. Resultater fra Storbritannia og USA antyder en intergenerasjonell utdanningselastisitet på mellom 0,2 og 0,45 (Black, Devereux og Salvanes, 2005). Sammenhengen gjelder også i Norge, og i følge Raaum (2003) er foreldrenes utdanning det viktigste familiekjennetegnet. Utdanningseffekten holder seg signifikant selv når husholdningsinntekt inkluderes. Chevalier et al. (2005) og Lauer (2003) sine resultater viser at mors utdanning er viktigst for barnet tidlig i utdanningsløpet. Etter grunnskolen har fars utdanning mer å si, men mors utdanning er fortsatt signifikant.

De fleste studier viser et signifikant og robust forhold mellom barna og mødrenes utdanning. Til tross for dette, fant Behrman og Rosenzweig (2001), en ikke-positiv effekt av mors utdanning da de kontrollerte for inntekt, og barnas og fedrenes evner målt i form av utdanning. De har derfor stilt spørsmål om effekten av mors utdanning egentlig viser effekten av hvem mor får barn med. Deres studie viser at et ekstra år utdanning til mor gir effekt i form av 0,4 år høyere utdanning hos far. Dette, sammen med den ikke-positive effekten av mors utdanning når man kontrollerer for andre ting, kan tyde på at den genetiske effekten av mors utdanning er større enn den kausale. Man har også funnet at korrelasjonen mellom barns

fødselsvekt, som er en viktig indikator for helse, og mors utdanning er positiv. (Chevalier et al., 2005).

Black, Devereux og Salvanes (2005) studerte sammenhengen mellom barn og foreldres utdanning. Funnene deres tyder på at den høye korrelasjonen mellom foreldres og barns utdanning først og fremst skyldes familiekarakteristika og arvede evner; ikke spillovereffekt av utdanning.

2.2.2 Inntekt

Familiens økonomiske situasjon setter restriksjoner for ressursene barnet har tilgang på. Fagmiljøene er derfor opptatt av hvilken effekt økonomi har på barns utdanningsdeltagelse. Foreldrenes inntekt er et vanlig mål på familiens økonomiske situasjon.

Familieinntekt kan påvirke barnas evneutvikling direkte og indirekte. Må man, helt eller delvis betale for utdanning av egen lomme, vil familiens økonomiske situasjon påvirke barnets utdanningsmuligheter. Kredittbegrensninger kan da føre til at fattige familier ikke har råd til å betale for høyere utdanning (Caneiro og Heckman, 2002). Settes dette i sammenheng med at kapitalmarkedene ikke er perfekte, betyr det at økonomiske skranker kan føre til at fattige underinvesterer i utdanning (Lauer, 2002). Dette kan også være tilfelle for risikoaverse personer (Hægland, 2003).

God økonomi gir flere valgmuligheter, og familieinntekt er sterkt korrelert med barnets livssyklus. Familier med høy inntekt i barnets ungdomsår, har med stor sannsynlighet også hatt høy inntekt da barnet var lite. Rike barn har tilgang til flere ressurser enn fattige barn, og det er derfor mer sannsynlig at rike har en livssituasjon som bedre fremmer utvikling av kognitive og ikke-kognitive evner (Caneiro og Heckman, 2002).

Inntekt kan også være en indikator på annet enn familiens økonomiske situasjon. Hvis inntekt avspeiler dyktighet, betyr det at høytlønnede har større evner enn lavtlønnede (Shea, 2000). Disse evnene kan påvirke barna både gjennom arv og mulighet til å tilby bedre oppvekstvilkår. Lav lønn kan i følge Raaum (2003) også skyldes at man har psykososiale problemer som gjør at man ikke fungerer godt i arbeidslivet. Dette er problemer som kan påvirke barnas oppvekstvilkår og evneutvikling negativt.

Resultater fra tidligere forskning

Det er liten tvil om at det er positiv samvariasjon i økonomisk status mellom generasjoner. Korrelasjonen mellom foreldreinntekt og barnets inntekt er sterk (Chevalier et al., 2005). I følge Shea (2000) er korrelasjon mellom fedre og sønners permanente inntekt 0,4. Inntekt har også positiv effekt på utdanningslengde. Barn av foreldre med høy inntekt går lenger på skolen – også lenger enn barn av foreldre med tilsvarende utdanning, men med lavere inntekt.

Samvariasjonen mellom fars inntekt og barnets utdanning viser at en forskjell på 100 000 kroner i lønn hos fedrene tilsvarer nesten ett skoleår for barnet. Inntektseffekten faller når man sammenligner barn i rike og fattige familier med lik utdanning, men det er likevel korrelasjon mellom foreldreinntekt og barnas utdanning (Raaum, 2003).

Til tross for dette er det ikke sikkert at det er pengene i seg selv som er av avgjørende betydning. En indikasjon på svak kausal sammenheng mellom rikdom og barneutdanning er at korrelasjonen mellom barnas utdanning og foreldreinntekt er større enn mellom barnas utdanning og familieformue. Dette kan tyde på at høy inntekt er korrelert med positive ikke-observerbare foreldreevner. En annen studie som tilbakeviser inntektseffekten er en forskningsrapport hvor man har funnet positivt forhold mellom inntekt og IQ til biologiske barn, men ikke til adopterte (Shea, 2000). Caneiro og Heckman (2002) har funnet at når man tar hensyn til individets prestasjoner på skolen, så forsvinner mye av inntektseffekten. De konkluderer derfor med at det er langsiktige familieforhold og ikke kortsiktige lånebegrensninger som spiller rolle for utdanning.

I studier hvor man ser hvordan kredittbegrensninger påvirker utdanningsvalg skiller man mellom øyeblikkelige og langsiktige kredittbegrensninger. Øyeblikkelige kredittbegrensninger vil si at familien har dårlig økonomi i perioden hvor barnet er ferdig med obligatorisk skole og skal starte på videregående, men ellers har klart seg greit. Langsiktig kredittbegrensning beskriver en situasjon hvor familien har hatt dårlig råd hele barnets livsløp. Ulike former for kredittbegrensninger påvirker utdanning forskjellig. I følge Dearden (2004) har øyeblikkelig kredittbegrensning litt å si, men lite sammenlignet med effekten av langsiktig kredittbegrensning.

For gruppen med lavest inntekt kan kredittbegrensninger ha noe å si. Raaum (2003) sier blant annet at økt inntektsnivå vil ha en positiv kausaleffekt på utdanningsnivået i fattige familier.

Dette finner man også i Sheas (2000) resultater. Han har funnet at for barn av fedre med mindre enn 12 års utdanning, vil økt pengebeholdning ha positiv effekt på utdanning. Da han kontrollerte for fars observerbare evner, hadde dette mye mindre effekt på MKM-estimatene til gruppen av fattige sammenlignet med det representative utvalget. Det er derfor mulig at fars utdanning og yrke er en svakere indikator på evne hos de fattigste. Effekten av kredittbegrensninger vil avhenge av utdanningstilbudet det offentlige gir, og i følge Black, Devereux og Salvanes (2005) har ikke kredittbegrensninger hatt avgjørende betydning for skolegangen til svakerestilte grupper i Norge.

2.2.3 Familie

Familien, og da særlig foreldrene, er de viktigste omsorgspersonene til barnet. Under avsnittet om foreldreutdanning og inntekt ble ressurser og tilrettelegging for utdanning fremlagt som viktige grunner til reproduksjon av sosioøkonomisk suksess. Bevisste foreldre kan, selv uten store midler og høy utdanning, legge sterke føringer for oppvekstmiljøet til barna. Raaum (2003) skriver for eksempel at familien former ungdoms preferanser for videre skolegang. Også Lauer (2002) mener at foreldrenes preferanser og verdier påvirker barnas syn på kostnader og avkastning av utdanning. Med bakgrunn i dette kan man si at så lenge familien ikke har sterkt begrensede ressurser så vil holdninger og kunnskap om hva som påvirker barns utdanning gi positiv effekt på oppvekstmiljø.

Størrelsen på familien er med på å styre hvilke ressurser barnet får tilgang på. Generelt kan man si at teori om vellykket utdanningsforløp, favoriserer høy andel voksenkontakt per barn. I følge Heckman (2005) er aleneforeldre ofte lavt utdannet og har dårlige evner. Fraværet av en forelder gjør at barnet får mindre voksenkontakt og stimuli. Dynamikken er den samme hvis to foreldre får flere barn. Raaum (2003) påpeker at dette vil redusere tid og ressurser per barn, siden barnetrygden ikke fullt ut kompenserer merkostnaden. I økonomenes terminologi innebærer flere barn at foreldrene har en lavere kapasitet til å investere ressurser i aktiviteter som styrker hvert enkelt barns utdanningskarriere. Mors alder tillegges også betydning. Jo eldre hun var da hun fikk barn jo høyere utdanning kan man anta at hun har.

Samspillet mellom søsken og plassering i flokken kan ha betydning for skolegang. Dette skjer for eksempel ved konkurranse om foreldrenes oppmerksomhet og etablering av rollemodeller. Barna som er født først får sannsynligvis mer oppmerksomhet i sine første leveår, dette er også perioden hvor kognitive evner dannes (Raaum, 2003).

Resultater fra tidligere forskning

Resultater fra tidligere forskningsprosjekter tyder på at oppvekstmiljøet har en avgjørende betydning for hvordan det går med barna. Det kan dokumenteres at barn fra risikofamilier⁶ har bedre karakterer, sjeldnere dropper ut fra skolen, har mindre spesialundervisning og lavere kriminell atferd enn kontrollgruppen, i tilfellene der myndighetene har grepet inn og sørget for økt stimuli på et tidlig tidspunkt (Heckman og Masterov, 2004).

Familiestrukturer ser også ut til å ha betydning. Samlivsbrudd er negativt for utdanningslengden, men effekten av skilsmisse er vanskelig å beregne (Raaum, 2003). Det er ikke gunstig for utdanningslengden å vokse opp med aleneforeldre. Andelen unge som dropper ut fra skolen er større i denne type familier enn i vanlige kjernefamilier. Var mor i tillegg tenåring da hun fikk barn er situasjonen enda mer dramatisk. Slike familier er forbundet med lav IQ, familieinntekt, emosjonell og intellektuell støtte (Heckman og Masterov, 2004).

Statistisk sett har antall søsken og plassering i flokken betydning for utdanningsforløp. Den generelle tendensen viser at barn i store familier har kortere utdanning enn barn fra små familier. Men sammenhengen mellom antall søsken og skoleår er ikke monoton - enebarn har kortere utdanning enn de med ett søsken, men deretter avtar utdanningslengden med antall søsken. Fellestrekkene mellom søsken er mange, og omkring 40 % av variasjonen i utdanningslengde er forklart ved kjennetegn ved familien, oppvekstmiljø eller arvelige egenskaper/forutsetninger som søsken deler. Korrelasjonen mellom tvillinger er betydelig høyere enn for søsken generelt – dette kan ha med at tvillinger blir oppdratt annerledes/likere enn andre søsken og opplever mye av de samme tingene (Raaum, 2003).

2.2.4 Nabolag og skole

Nabolageffekten er et samlemål på innflytelsen fra venner og skolekamerater i nærmiljøet, sosiale normer og holdninger som er dominerende der en vokser opp, samt lokale institusjoner som skoler/barnehager og fysisk miljø (Raaum, 2003). Nabolaget er med på å sette sosiale standarder som påvirker utdanning. Læringsmiljøet i en klasse avhenger i stor grad av hvilke holdninger toneangivende elever har til skole. Det er derfor grunn til å anta at å gå i en klasse

⁶ Med risikofamilie menes her at familien er fattig og at mor er alene med ansvaret for barna.

med ressurssterke elever som er interessert i læring vil skape et fordelaktig læringsmiljø. Kvaliteten på skolen antas også å ha betydning. Jo bedre lærere, fasiliteter og skolemateriell, jo flinkere elever.

Resultater fra tidligere forskning

Det er uenighet om hvor mye kvaliteten på skolen påvirker barnets kognitive evner. Undersøkelser tyder på at mer ressurser til skolen ikke gir veldig mye bedre resultater for elevene. Den reelle effekten av økte skolemidler er vanskelig å bestemme siden familieinput er utelatt fra analysen, men mønsteret tyder på at bedre skole ikke gir bedre elever (Todd og Wolpin, 2003). Lauer (2002) har sammenlignet effekt av skolesystemet i Frankrike og Tyskland, og var overrasket over hvor små forskjeller hun fant mellom relativt forskjellige utdanningssystemer.

Nabolag og skolemiljø har likevel en viss betydning. I USA har man gjort studier hvor familier har blitt flyttet fra dårlige til gode nabolag. Disse har gjort det merkbart bedre på skolen. Om dette skyldes nabolag, skole eller inntekt (familiene fikk hjelp til leieutgiftene), er vanskelig å si (Chevalier et al., 2005). For norske barn født i perioden 1945-1955 er nabolagseffekten 0,11 for kvinner og 0,12 for menn, men det er grunn til å tro at anslagene overvurderer nabolagseffekten. For det første avspeiler de at nabolagene består av like familier. Sorteringen av familier etter nabolag innebærer at nabounger ofte har svært likt hjemlig oppvekstmiljø. For det andre er det nærliggende å tro at fordelaktig nabolag går sammen med fordelaktig familiebakgrunn. Nabokorrelasjonen i Norge blir svakere over tid. Dette kan ha sammenheng med inntektsutjevningen på 60- og 70-tallet, og at betydningen av familiebakgrunn på utdanningsdeltagelsen svekkes (Raaum, 2003).

2.2.5 Avstand fra hjemsted til høyere utdanningsinstitusjon

Distansen mellom hjemsted og høyere utdanningsinstitusjon kan påvirke deltagelsen til høyere utdanning direkte gjennom finansielle og emosjonelle kostnader, og indirekte gjennom nabolagseffekter. Det er naturlig å tenke seg at kostnadene ved å studere øker jo lenger kilometeravstanden er fra hjemsted til studiested. Aamodt peker blant annet på at det kan være store økonomiske gevinster å hente på å bo hjemme hos foreldrene mens man studerer. Bor foreldrene i pendleravstand til utdanningsinstitusjonen, kan barna slippe å betale for kost og losji. Sammenlignet med de som må flytte hjemmefra for å studere kan dette gi store økonomiske fordeler Aamodt (1982). Det kan også være emosjonelle kostnader knyttet til å

flytte fra hjemstedet. Det er ikke sikkert at man er villig til å gi slipp på sosiale nettverk av familie og venner for å begynne på høyskole eller universitet. En tredje årsak kan være at personer som er bosatt langt unna høyere utdanningsinstitusjoner ikke ser fordelene ved å ta utdanning. Denne teorien baserer seg på en antakelse om at færre innbyggere i rurale strøk har høyere utdanning, og at dette slår negativt ut på den generelle holdningen til utdanning i området. Effekten av avstand mellom hjemsted og undervisningssted blandes på denne måten sammen med nabolageeffekten (Frenette, 2002).

Resultater fra tidligere forskning

Resultatene fra tidligere forskning er entydige. Jo lenger borte fra høyere utdanningsinstitusjonene du bor, jo mindre sannsynlig er det at du tar høyere utdanning. Frenette (2002) finner for eksempel at individer bosatt mer enn 40 km borte fra universitetet bare har 75 % sannsynlighet for å ta universitetsutdanning, sammenlignet med kvantitativt like individer bosatt mindre enn 40 km vekk fra universitetet. De samme resultatene kommer Card (1993) fram til. Menn bosatt i områder nær fireårige college har et signifikant høyere nivå av utdanning og inntekt. Effekten av avstand er derimot ikke lik for alle gruppene. Det ser ut til at de gruppene det er mest sannsynlig at vil ta høyere utdanning i utgangspunktet, individer fra familier med høy inntekt eller at foreldrene har høyere utdanning, blir mindre påvirket av avstand. Frenette (2002) fant for eksempel at barn fra høyinntektsfamilier har 1,9 ganger større sannsynlighet for å ta universitetsutdanning enn barn fra lavinntektsfamilier når begge familiene er bosatt innenfor 40-kilometersgrensen. Tilsvarende tall for familiene bosatt lenger vekk er 5,1. Når man tar hensyn til inntekt, er effekten av avstand den samme for barn av foreldre med og uten høyere utdanning. Den positive effekten av å være kvinne reduseres kraftig når avstand inkluderes. De mest negative konsekvensene av avstand til utdanningsinstitusjon er for barn av lavinntektsforeldre og kvinner. Dette stemmer også overens med Card (1993) sine funn. Han finner at effekten av nære college er størst for menn med lavest antatt skoledeltagelse, altså de som kommer fra relativt fattige familiebakgrunner.

2.3 Svakheter ved tidligere forskning

Kritikken som er rettet mot tidligere forskning kan deles i to. En del angriper metodikken forskningsgruppene har brukt. Dette dreier seg i hovedsak om problemstillinger knyttet til feilspesifisering av modeller og måleproblemer. Den andre delen retter seg mot hvordan

resultatene benyttes, både når det gjelder sammenligning mellom studier og konklusjoner som trekkes på bakgrunn av funnene.

2.3.1 Modelladopsjon uten teoretisk fundament

Det teoretiske fundamentet bak de empiriske studiene er tidvis dårlig. Dette skyldes blant annet at modellene adopteres ukritisk, eller at studiene baseres på antagelser og ikke faste teorier. Effekten av dette kan være at modellene feilspesifiseres eller at resultatene blir vanskelig å tolke og/eller sammenligne.

Det er lite konsensus om hvilke faktorer som er viktige for evneutviklingen. Todd og Wolpin (2003) peker på at studier som baserer seg på samme datagrunnlag får vidt forskjellige resultater. De mener dette skyldes at de statistiske modellene som benyttes er feilspesifiserte. Modellene mislykkes derfor med å ta hensyn til viktige faktorer som bestemmer prestasjon. I tillegg er ligningene som benyttes til å finne estimatene ofte adoptert uten særlig rettferdiggjøring. Dette gjør at det er vanskelig å vite om antagelsene som ligger til grunn for arbeidet er troverdig. Lauer (2002) viser også til at det kan være vanskelig å tolke estimerte parametere når litteraturen begrenser seg til å etablere en statistisk link mellom sosialt opphav og utdanningsmessig prestasjon, uten å avhenge av en teoretisk modell.

2.3.2 Måleproblemer, stedfortredende variabler og instrumentvariabler

Evner er vanskelig å måle, og dette innebærer at sannsynligheten for at det oppstår målefeil eller at data mangler, øker. Slike problemer gjør resultatene mindre presise og i verste fall ubrukelige. For å kompensere for svakt datamateriale benyttes ofte stedfortredende (proxy) variabler og instrumentvariabler (IV). I teorien er dette effektive og gode hjelpemidler, men i praksis er bildet mer komplisert. Vesentlige deler av kritikken som har kommet mot forskning på utvikling av evner, og sosial mobilitet i forhold til utdanning, dreier seg om bruk av dårlige proxyer og instrumentvariabler.

Proxyvariabler

Proxyvariabler brukes når den egentlige forklaringsvariabelen ikke lar seg måle på en presis måte. En proxy er en observert variabel som er relatert, men ikke identisk til den uobserverte forklaringsvariabelen (Wooldridge, 2003). Denne/disse variablene inkluderes fordi man antar

at proxyene vil begrense den utelatte variabelskjevheten som oppstår fordi man ikke har tilgang på data om den variabelen som virkelig påvirker regressanden.

Under gitte forutsetninger som tilfeldig målefeil i proxyen og null målefeil i de andre inputvariablene, kan bruk av proxy redusere utelatt variabelskjevhet (Todd og Wolpin, 2003). I praksis er det derimot vanskelig å finne proxyer som er sterkt nok korrelert med variabelen vi mangler (Wooldrige, 2003).

Dårlige proxyer kan gjøre modellkoeffisientene enda skjevare og tolkningen vanskeligere. Et eksempel på det første er når utgift per elev inkluderes som proxy for manglende data om andre skoleinputs. Skoler med identisk utgiftsnivå per elev, men med forskjellig klassestørrelse, vil her komme likt ut selv om skolen med færrest elever per klasse bruker mindre på andre uobserverte inputs, for eksempel lærere. Den målte effekten av klassestørrelse på prestasjon, avhengig av utgifter per elev, er netto av effekten av uobserverte inputs. Hvis klassestørrelse og uobserverte inputs ikke er korrelert, slik at utelatt variabelskjevhet ville vært null i modellen uten proxy, så vil man ved å inkludere utgiftene i regresjonen være sikker på at bevegelser i klassestørrelse blir blandet sammen med bevegelse i uobserverte inputs. Å inkludere en proxy i dette tilfellet vil føre til større skjevhet (Todd og Wolpin, 2003).

På samme måten kan inklusjon av proxyer gjøre det vanskeligere å tolke resultatene. Dette ser vi blant annet når inntekt benyttes som proxy for manglende data om familiekarakteristika som kan påvirke barnas utdanningsprestasjoner. Antagelsen bak å inkludere denne variabelen er at familier med høy inntekt kjøper flere ting som kan være med på å utvikle barnas evner, for eksempel bøker. Dersom inntekten holdes konstant, vil kjøp av flere bøker innebære en reduksjon i kjøp av andre varer som også kan påvirke evneutviklingen – eksempelvis stimulerende leker. Inklusjon av proxyen vanskeliggjør på denne måten tolkningen, fordi effekten av å øke antall bøker blandes sammen med effekten av reduksjon av andre goder (Todd og Wolpin, 2003).

Instrumentvariabler

Instrumentvariabler er en annen måte å hankses med at man ikke kan måle alle variablene som antas å ha effekt på regressanden. En IV er en variabel som ikke inngår i

regresjonsligningen, men som er delvis korrelert med den endogene forklaringsvariabelen, men ikke med feilledet (Wooldrige, 2003). Metoden gjør at man kan finne estimater selv om ikke alle variablene lar seg måle. Manglende mulighet til å observere evner er en vanlig årsak til at instrumentvariabler benyttes i denne type litteratur. Problemet med instrumentvariabler er at metoden gir skjeve estimater hvis ikke instrumentene er gode nok.

Caneiro og Heckman (2002) stiller seg tvilende til gyldigheten av vanlige instrumenter som antall søsken, mors og fars utdanning, avstand til høyere utdanning og lokal arbeidsledighet. De har hatt tilgang til en undersøkelse som inneholder data på AFQT, som er et mål på evner. Dette har de benyttet til å teste gyldigheten til forskjellige typer instrumenter ved å estimere korrelasjon mellom instrument og evner. Konklusjonen som trekkes er at eksisterende litteratur er infisert av dårlige instrumentvariabler - instrumentene er enten korrelert med både utdanning og evner, eller ukorrelert med evner. Todd og Wolpin (2003) påpeker også at instrumentvariablene ikke er gyldige hvis utelatte inputs ikke er uavhengig til de inkluderte variablene. Sannsynligheten for at utelatte variabler som reflekterer valg, og inkluderte variabler er korrelert med instrumentene er like stor. Dette innebærer at et hvert instrument som har kraft, også ikke vil være gyldig.

2.3.3 Manglende sammenlignbarhet

En stor del av anvendelsen av empiriske studier består i å sammenligne resultater. Dette gjøres for å kartlegge forskjeller mellom grupper eller utvikling over tid. Men ikke alle studier kan sammenlignes. I følge Todd og Wolpin (2003) gjelder dette spesielt sammenligning av resultater fra eksperimentelle og ikke-eksperimentelle undersøkelser, fordi forskjellige typer undersøkelser ofte belyser forskjellige aspekter ved en problemstilling.

Parametere som estimeres i eksperimentelle studier avviker typisk fra de som estimeres i ikke-eksperimentelle studier, og en type bevis er ikke et substitutt for et annet. Effekten av skolekvalitet på elevenes prestasjoner er et eksempel på et område hvor det finnes begge typer forskning. Studier som baserer seg på eksperimentelle data brukes generelt til å identifisere effekt av politikk, mens ikke-eksperimentelle data vanligvis identifiserer parametere i utdanningsproduksjonsfunksjonen. Parameterne sammenlignes ofte som om de er fra samme type studie. Dette er ikke bra siden de to typene studier svarer på forskjellige spørsmål, og det derfor ikke er noen grunn til å forvente at de ulike estimatene skal matche hverandre (Todd og Wolpin, 2003).

2.3.4 Forveksling av samvariasjon og kausalitet

Kvantitative studier gir korrelasjon mellom ulike variabler, men resultatene inneholder ikke kunnskap om hvordan variablene henger sammen. Samvariasjonen mellom ulike familiekjennetegn og utdanningskarrierer representerer derfor ikke nødvendigvis kausale effekter (Raaum, 2003). Noen ganger forveksles likevel korrelasjon mellom variabler med kausalitet, dette kan være problematisk. Dersom konklusjonene som trekkes ikke er korrekte, gjør det kunnskapen om emnet mindre presis. Dette kan ha konsekvenser både for utarbeiding av politikk og videre forskning.

Ofte er det nødvendig med innsikt om årsakssammenhenger for å innrette utdanningspolitikken slik at mulighetene som barn og unge tilbys i mindre grad bestemmes av hvilken familie en har fått tildelt (Raaum, 2003). I den grad utarbeiding av politikk baseres på kunnskap om kausale sammenhenger, vil feilaktige slutninger føre til svekket mulighet til å påvirke samfunnet fra myndighetenes side.

Misvisende konklusjoner preger også videre forskning gjennom følgefeil. Shea (2000) kommer med eksempel på denne type problemstilling. Et viktig resultat fra tidligere forskning har vært at høyere foreldreinntekt gjør at barna tar lengre utdanning. Det har vært antatt at det er et kausalt forhold mellom disse variablene. Shea (2000) stiller spørsmål om høy korrelasjon mellom foreldreinntekt og barnas utdanning egentlig skyldes sammenhengen mellom evner og høy inntekt. I den grad det er sammenheng mellom evner og inntekt, er det naturlig å tenke seg at det er barnas arvede evner, mer enn foreldrenes inntekt, som gjør at de tar lang utdanning. Sheas (2000) resultater er generelt ikke støttende til modellen om at foreldrenes inntekt spiller rolle for barnas utdanning. Dette er et eksempel på at konsensus blant forskere om årsakssammenhenger, ikke nødvendigvis er korrekte.

KAPITTEL 3: DATABESKRIVELSE

Dette kapitlet gir en oversikt over datamaterialet som benyttes i oppgaven. For å besvare problemstillingen var det nødvendig med noen justeringer av dataene. Kapitlet beskriver derfor hvilke endringer som er gjort, og begrunnelsen for disse. Den deskriptive delen av kapitlet består av mer utfyllende informasjon om variablene som benyttes i regresjonene. Her inngår også en del sammenhenger som ikke direkte er med i det empiriske arbeidet, men som likevel benyttes i analysen.

3.1 Datasett

Datasettet stammer fra Statistisk sentralbyrå og inneholder informasjon om kohorten født i 1968. Dataene er organisert som tverrsnitt, og inkluderer informasjon om antall barn i familien, antall individer som forsørgeres av inntektene i familien, barnas lønnsinntekt fra 1990 til 2000, nasjonalitet, fødeland, antall år med utdanning, kjønn, sivilstatus, kommune- og fylkestilhørighet til barnet i 2000, kommunetilhørighet for mor i 1960, høyeste fullførte utdanning til foreldre og barn, foreldrenes alder, og fars og mors pensjonsgivende inntekt i perioden 1967-1999. SSB har koblet sammen data fra ulike offentlige registre. Individene som beskrives i datasettet har derfor ingen direkte påvirkning på informasjonen som er tilgjengelig om dem.

Formålet med denne oppgaven er å beskrive klassiske sammenhenger og dynamikker når det gjelder utdanning i Norge. Individer som representerer brudd på norsk "normalitet" kuttet derfor ut. Det vil si individer som kommer fra familier hvor det er mer enn seks barn i familien, hvor antall familiemedlemmer som forsørgeres overstiger åtte, hvis mors alder er under 16 eller over 46 når barnet er ett år, tilsvarende hvis far er under 19 eller over 55 år. Jeg har også utelukket utenlandske statsborgere og barn født i andre land. Det samme gjelder individer med mindre enn ni års skolegang, som samsvarer med obligatorisk skolegang for kohorten. I utgangspunktet består datamaterialet av 71 327 individer, men på grunn av manglende observasjoner eller dårlig representativitet utelukkes ca 30 % av disse.⁷

Fokus i oppgaven er hvilken effekt geografisk bosted har på utdanningsdeltagelsen. Av flere grunner er det komplisert å kartlegge distansen mellom hjemsted og nærmeste høyere utdanningsinstitusjon for en hel kohort.⁸ Av praktiske årsaker baserer analysen seg derfor på

⁷ Dette materialet brukes til å sjekke representativiteten for resultatene.

⁸ I avsnitt 3.1.4 går jeg nærmere inn på denne diskusjonen.

individer med fylkestilhørighet i Hordaland og Sør-Trøndelag. Nettoutvalget som benyttes består av 8480 individer.

3.1.1 Utdanning

Datamaterialet inneholder informasjon om barnas og foreldrenes høyeste fullførte utdanning. Dette er organisert ved at hver enkelt utdanning har en spesifikk kode. Ved fullført utdanning knyttes den aktuelle koden til individet. Første siffer i koden angir utdanningens nivå. Det vil si at alle utdanninger på for eksempel forskernivå begynner med 8, uavhengig av hvilket fagfelt det forskes på. Kodesystemet som benyttes, Norsk standard for utdanningsgruppering (NUS), ble som følge av store endringer i utdanningssystemet erstattet med et nytt system, NUS2000, i år 2000. Foreldrenes utdanning er registrert i 1980, altså etter gammel ordning, mens barnas utdanning er registrert med det nye kodesystemet i 2000. De to systemene er ikke direkte kompatible.

Hovedendringene mellom systemene består i at man har gått fra en tre-sifferinndeling av utdanning til seks-sifferinndeling. Første siffer i hvert av systemene angir i begge tilfeller nivået på utdanningen, men hvilket nivå sifferet indikerer er ikke enhetlig. Standarden fra 2000 har tre nivåer for videregående eller påbygging til videregående utdanning, mens den gamle standarden bare har to. Forskjellen innebærer blant annet at forkurs som verken gir vektall/studiepoeng, eller er en integrert del av en universitets- eller høyskoleutdanning etter gammel standard regnes under gymnasnivå II (som ellers innebærer avsluttende videregående utdanning), mens disse etter ny standard har et eget nivå kalt påbygging til videregående utdanning. På høyskole- og universitetsnivå er situasjonen omvendt. Gammel standard opererer med tre nivåer, mens ny standard har to. Det har derfor vært nødvendig å lage en egen nivå-inndeling som tar hensyn til begge systemene. Denne består i 8 nivåer. Fra 0 som innebærer ingen registrert utdanning eller førskoleutdanning, til 7 som er forskerutdanning. For foreldrene har jeg derfor slått sammen nivå 5 og 6 som begge indikerte lavere grads studier på høyskole- eller universitet, og for barna er VK II og nivået som NUS 2000 kaller påbygging til videregående utdanning slått sammen.

Utdanning klassifiseres på tre ulike måter i denne oppgaven. I tillegg til nivåinndelingen som er beskrevet overfor, vurderes utdanning også ut i fra høyeste antall fullførte skoleår, og som

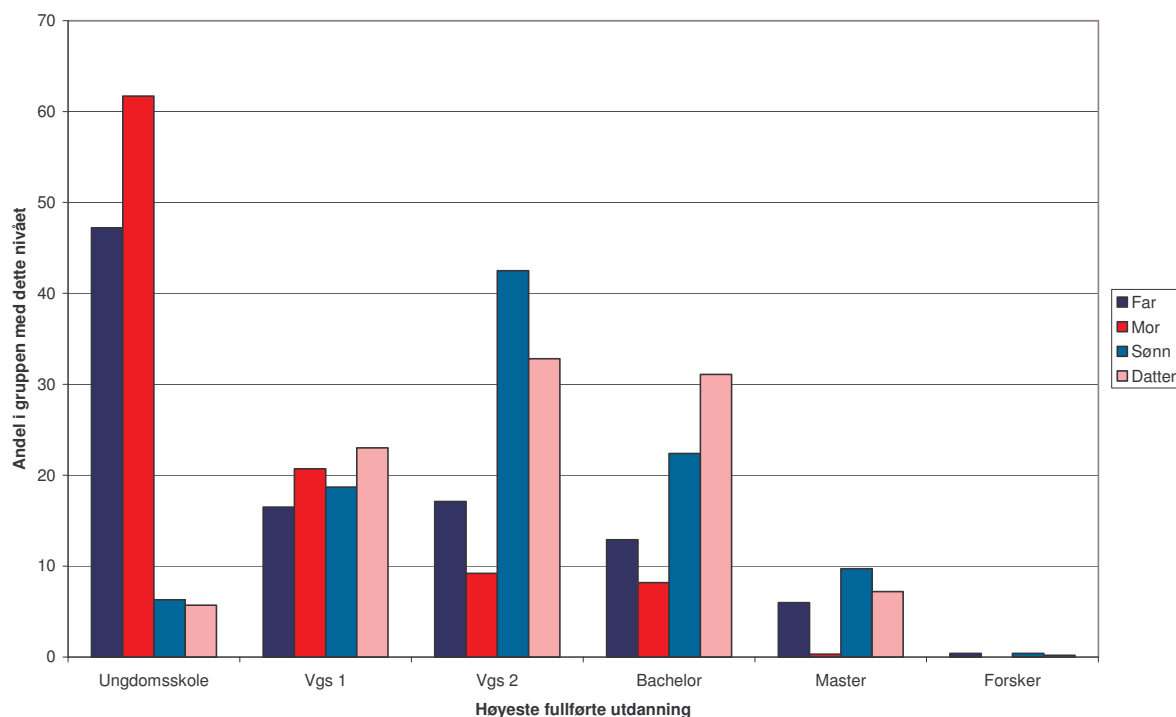
en binær variabel som identifiserer om individet har høyere utdanning eller ikke. Størrelsene er tilnærmet kompatible⁹ og konverteringen mellom nivåene vises i tabell 3.3

Tabell 3.1 Kvalitativt innhold av utdanningsnivå.

Nivå	Kvalitativt innhold av nivå	Antall år utdanning
0	Ingen utdanning	0
1	Barneskole	6
2	Ungdomsskole	9
3	Videregående 1	11
4	Videregående 2	12
5	Universitet/ Høyskole 3 år, Bachelor	15
6	Universitet/ Høyskole 5 år, Master	17
7	Forskerutdanning	17+

Utdanningsnivået til foreldre og barn i utvalget presenteres grafisk i et histogram.

Figur 3.1 Høyeste fullførte utdanningsnivå for foreldre og barn



⁹ Antall år med skolegang stemmer ikke nødvendigvis overens med utdanningsnivåene for alle individene. Et eksempel på dette er Cand.mag-graden fra universitetet. Denne er på fire år, men regnes som nivå 5 lavere grad på universitet/høyskole, som tilsvarer tre års høyere utdanning.

Vi ser av tabellen at utdanningsnivået i befolkningen har økt mye fra foreldregenerasjonen til barna. Dette gjelder spesielt fullført videregående skole og bachelornivå på høyskole/universitet. 13 % av foreldrene har videregående 2 som høyeste fullførte utdanning. Hos barna har dette økt til nesten 38 %. Tilsvarende tall for bachelorgraden er 11 %, mot 27 %.

Økning i utdanningsnivå gjelder spesielt kvinner. Andelen kvinner i 68-kohorten med høyere utdanning er 30 % større en tilsvarende andel hos mødrene.¹⁰ Veksten i høyere utdanning er også markant sterkere for kvinner enn for menn, og i dette utvalget har flere kvinner enn menn høyere utdanning. Majoriteten av de som fullfører de høyeste utdanningene, som mastergrad og doktorgrad, er likevel menn.

Sammenlikner man 68-kohorten, og foreldrene kan man trekke slutningen om at det ikke bare har skjedd en endring i antall studenter, men også en markant endring i *hvem* som tar høyere utdanning. Før var brorparten av studentene menn, nå er det flest kvinner. Det ser også ut til at den første store utdanningsstoppen, der mange faller fra, kommer på videregående nivå for denne kohorten. For foreldrene skjedde dette på ungdomsskolen.

3.1.2 Inntekt

Datasettet inneholder informasjon om foreldrenes og barnas pensjonsgivende inntekt. Inntekt omfatter her lønnsinntekt, inntekt fra selvstendig virksomhet, arbeidsledighetstrygd, sykepenges og uføretrygd. Inntektene er oppgitt brutto. Dette innebærer blant annet at skatt ikke er trukket fra. Kapitalinntekter og renteinntekter er heller ikke inkludert. Et resultat av dette er at de registrerte lønnsinntektene ikke nødvendigvis representerer familiens reelle disponible pengebeholdning. Mors og fars inntekt er oppgitt separat. For å få familiens samlede lønnsinntekt i det aktuelle året, har jeg valgt å slå mors og fars inntekter sammen til en ny variabel kalt familieinntekt.

Det kan tenkes at effekten av økonomiske ressurser på evneutviklingen er forskjellig i ulike perioder av oppveksten. For å fange opp dette deles familieinntektene inn i ulike perioder.

¹⁰ En liten del av denne veksten kan forklares ut i fra at man i 1981 la om kategoriene for hva som regnes som høyere utdanning. Blant annet gikk utdanninger som sykepleier, førskolelærer og ingeniør over fra å regnes som fagutdanning til å få høyskolestatus. Dette er yrker som også foreldregenerasjonen hadde, men som da ble klassifisert lavere. Omleggingen favnet i større grad om typiske kvinneutdanninger (SSB og Vabø og Aamodt , 2005).

Variablene inneholder familiens gjennomsnittsinntekt i perioden, og periodene er inndelt slik at de er sammenfallende med kohortens utdanningsforløp.

Tabell 3.2 Oversikt over inntektsvariablene som benyttes

Inntekt (0-19)	Permanentinntekt, familiens gjennomsnittsinntekt i perioden 1968-1987
Inntekt (0-6)	Pre skole (0-6 år). Familiens gjennomsnittsinntekt i perioden 1968-1974
Inntekt (7-12)	Barneskole (7-12 år). Familiens gjennomsnittsinntekt i perioden 1975-1980
Inntekt (13-15)	Ungdomsskole (13-15 år). Familiens gjennomsnittsinntekt i perioden 1981-1983
Inntekt (16-19)	Videregående skole (16-19 år). Familiens gjennomsnittsinntekt i perioden 1984-1987

KPI-justering

Inntektene er oppgitt nominelt. Skal de avspeile reell kjøpekraft og kunne sammenlignes mellom ulike tidsperioder; må de inflasjonsjusteres. Alle inntekter er derfor deflatert til 1998-kroner etter tabell fra SSB.

Diskontering

Realverdien av en krone i 1967 og en krone i 1999 er den samme. Likevel er det ikke likegyldig i hvilken periode man mottar kronen. En krone mottatt i 1967 kan investeres, og verdien av den aktuelle kronen vil være langt høyere i 1999 enn en tilsvarende krone mottatt i 1999. Dette gjelder også for foreldrenes inntekt. Høy realinntekt tidlig i perioden vil være gunstigere enn tilsvarende høy realinntekt på et senere tidspunkt. For å avspeile reell verdi, bør inntektene derfor diskonteres før de benyttes i regresjonene.¹¹ Foreldrenes inntekt diskonteres derfor med 3 % med 1967 som utgangspunkt.

Tabell 3.3 Deflaterte og diskonterte familieinntekter, utvalgte år.

År	Deflatert, 1998kr	Diskontert, 1967kr
68	224302	217770
73	281298	235583
78	335759	242559
83	354881	221150
88	376676	190119
93	410086	202482
98	432792	173111

Tabellen viser oversikt over inntektsutviklingen for deflaterte og diskontert inntekter i utvalgte år. Deflaterte inntekter svinger mellom i underkant av 225 000 kr og 433 000 kr. Den

¹¹ Dette er gjort i Caneiro og Heckman (2003).

generelle tendensen er at familiens inntekter øker jevnt og trutt. Dette stemmer godt overens med vanlig lønnsutvikling.

Siden inntekt er oppgitt på både nominell, deflatert og diskontert form er det hensiktsmessig at resultatene kommer som forholdsmessige endringer, i stedet som for reelle størrelser. Inntektene om benyttes i regresjonene er derfor oppgitt logaritmisk.

3.1.3 Familiekarakteristika

Søskeninformasjonen i datasettet stammer fra 1970. For mange av barna i 68-kohorten vil ikke antall søsken som er registrert, samsvare med det reelle antall søsken de har, da søsken født etter 1970 ikke er tatt med. Søskenantallet kan likevel gi innblikk i livssituasjonen til barnet i dets første leveår når de kognitive evnen utvikles.

Tabell 3.4 Prosentvis fordeling over antall barn i familiene i 1970.

Antall barn	
1	25,5
2	39,5
3	21,8
4	9,5
5	3,4
6	0,3

Vi ser at 25,5 % av barna er enebarn, nesten 40 % har ett søsken, 22 % har to søsken, 9,5 % tre søsken, mens fire og fem søsken er mer uvanlig. Foreldrenes gjennomsnittsalder i 1970 var 28,2 år for mødrene og 31,1 år for fedrene.

3.1.4 Avstand mellom hjemsted og nærmeste høyere utdanningsinstitusjon

En måte å måle effekten av avstand mellom hjemsted og undervisningssted på deltagelse i høyere utdanning er å inkludere avstand som en variabel i regresjonen.¹² I praksis kan dette gjøres ved å lage en avstandsmatrise som beskriver distanse i mil eller reisetid, mellom kommunesentrum og nærmeste høyere utdanningsinstitusjon, og knytte dette opp mot hvert enkelt individ.¹³

¹² Dette er blant annet gjort av Frenette (2002 og 2003) og Card (1993).

Datasettet som benyttes inneholder ikke informasjon om barnas oppvekstkommune, men vi har kjennskap til kohortens bostedskommune i 2000 og mors bostedskommune i 1960. Siden individer med høy utdanning i større grad enn andre kan være avhengig av å bo i store kommuner for å få jobb, gir ikke nødvendigvis den registrerte bostedskommunen i 2000 et godt bilde av hvor individet har vokst opp. Mors bostedskommune i 1960 er heller ikke en perfekt indikator på hvor barnet har vokst opp, men det er likevel sannsynlig at dette er et bedre mål enn barnas bostedskommune i 2000.¹⁴ Mors bostedskommune i 1960 brukes derfor som utgangspunkt når avstand mellom hjemsted og utdanningssted skal beregnes.

Et annet problem ved å estimere effekten avstand har på deltagelsen i høyere utdanning, er at individene ikke bare tar høyere utdanning, de tar spesifikke høyere utdanninger. For de fleste er det ikke likegyldig om man blir siviløkonom eller sykepleier. Reell avstand mellom hjemsted og nærmeste høyere undervisningssted varierer derfor fra individ til individ. Slike forhold er vanskelig å kartlegge for en hel kohort. På grunn av dette har jeg valgt å fokusere på individer fra med fylkestilhørighet i Hordaland og Sør-Trøndelag.¹⁵ Dette er fylker hvor Bergen og Trondheim fremstår som naturlige utdanningsentrum. Byene har både universitet og et bredt spekter av høyskoleutdanninger, og det øvrige utdanningstilbudet i fylkene er nesten ikke-eksisterende.¹⁶ På grunn av dette, vil Bergen og Trondheim for svært mange representere det nærmeste reelle utdanningsstedet. Avstandsmatrisen baserer seg derfor på avstand mellom mors bostedskommune i 1960 og Bergen for individer fra Hordaland, og avstand til Trondheim for barna fra Sør-Trøndelag.

Både transportmidler og veier har blitt bedre siden majoriteten av individene tok fatt på sin høyere utdanning. Dette innebærer at reisetid, som er det faktiske hinderet ved å bo langt unna utdanningsinstitusjonene, har blitt forkortet. Det er ikke sikkert at forbedringene gir seg utslag på samme måte for alle reiseruter, men denne forenklingen gjør jeg likevel når avstandsmatrisen utarbeides. Kilometeravstanden og kjøretid mellom kommunesenter og

¹³ Avstandsvariablene Frenette (2003) benyttet var dummyer som indikerte om universitetet lå i pendleravstand. I dette arbeidet vil avstand inkluderes som en kontinuerlig variabel.

¹⁴ Svakheten ved å benytte dette som utgangspunkt for avstandsmatrisen diskuteres videre i avsnitt 6.2.

¹⁵ Siden avstandsmatrisen baserer seg på mors kommune i 1960, innebærer fylkestilhørighet her at mor bodde i Hordaland eller Sør-Trøndelag i 1960.

¹⁶ Den eneste høyere utdanningsinstitusjonen i fylkene som er lokalisert utenfor byene er Høgskolen Stord-Haugesund.

Bergen/Trondheim har jeg funnet ved hjelp av karttjenesten til Gule Sider. Disse baserer seg på dagens kommunikasjonsmuligheter. Måleenhet er mil.

3.2 Deskriptiv statistikk

I den deskriptive delen oppsummeres informasjon om dataene som benyttes i det empiriske arbeidet. I tillegg studeres sammenhenger mellom utvalgets høyeste fullførte utdanning og foreldreutdanning, inntekt og avstand.

3.2.1 Deskriptiv statistikk om variablene som benyttes i regresjoene

Deskriptiv statistikk om variablene som benyttes i regresjonene vises i tabell 3.4 og 3.5. For å fange opp kjønnsforskjeller separer jeg menn og kvinner.

Tabell 3.5 Deskriptiv statistikk, menn

Variabel	Obs	Gj.snitt	St.avvik	Min	Maks
Antall år utdanning	4282	12,69	2,35	9	21
Utdanningsnivå	4282	4,12	1,04	2	7
Høyere utdanning	4282	0,32	0,47	0	1
Inntekt (0-19)	4282	219935,5	69710,59	4730,1	628345,4
Inntekt (0-6)	4278	193679,2	64142,71	119,05	525288,9
Inntekt (7-12)	4282	227306	75578,24	0	684964,7
Inntekt (13-15)	4281	251645,7	94705,91	0	1184773
Inntekt (16-19)	4282	250027,9	112232,2	0	1043270
Mors utdanningsnivå	4282	2,66	0,98	2	7
Fars utdanningsnivå	4282	3,14	1,31	2	7
Antall barn i familien	4282	2,27	1,08	1	6
Mors alder	4282	28,25	2,45	18	46
Fars alder	4282	31,15	6,36	19	55
Avstand	4282	3,97	5,21	0	20,5

Tabell 3.6 Deskriptiv statistikk, kvinner

Variabel	Obs	Gj.snitt	St.avvik	Min	Maks
Antall år utdanning	4198	12,79	2,35	9	20
Utdanningsnivå	4198	4,12	1,03	2	7
Høyere utdanning	4198	0,39	0,49	0	1
Inntekt (0-19)	4198	221466,3	70064,77	4531,7	857254,4
Inntekt (0-6)	4195	192740	64405,85	0	568819,4
Inntekt (7-12)	4198	226349,8	73698,49	0	825685,4
Inntekt (13-15)	4196	253949,9	99161,92	0	1387577
Inntekt (16-19)	4198	256214,2	111171,3	0	1227336
Mors utdanningsnivå	4198	2,63	0,96	2	6
Fars utdanningsnivå	4198	3,16	1,33	2	7
Antall barn i familien	4198	2,27	1,05	1	6
Mors alder	4198	28,22	5,37	18	46
Fars alder	4198	31,19	6,27	20	55
Avstand	4198	3,88	5,15	0	20,5

Tabellen viser at gjennomsnittelig antall år utdanning for mennene er 12,69 år. Dette samsvarer med et gjennomsnittelig utdanningsnivå i overkant av 4, som tilsvarer litt over fullført videregående skole. Kvinner har i gjennomsnitt 0,1 år mer utdanning. Forskjellene mellom kjønnene kommer klarere frem når man studerer andel med høyere utdanning i utvalget. 32 % av mennene har høyere utdanning, mot 39 % av kvinnene.

Som ventet er det ikke store forskjeller i familiekarakteristika. Gjennomsnittsinntekter, foreldres utdanningsnivå, alder og antall søsken ligger omtrent på samme nivå. Dette gjelder også for avstand. Gjennomsnittsavstanden til Bergen/Trondheim ligger litt i underkant av fire mil. Den korte gjennomsnittsavstanden skyldes at nesten halvparten av utvalget bor i byene.

3.2.1 Sammenheng mellom familiekarakteristika og barnas utdanning

Det er en politisk målsetting at barns utdanning i minst mulig grad skal avhenge av foreldrenes sosioøkonomiske situasjon eller geografisk bosted.¹⁷ Det er derfor naturlig å evaluere hvilke sammenhenger som finnes mellom barns utdanning og familiekarakteristika. Jeg sammenligner derfor utdanningen til barn fra ulike sosioøkonomiske grupper.

Barna ble delt inn i ulike grupper på følgende måte. Jeg skilte barna som har foreldre med høyere utdanning fra barna hvor foreldrene ikke har høyere utdanning. Utvalget ble delt i tre etter familiens inntekt. For å fange opp effekten av avstand ble barna delt inn i grupper etter om de bodde i byene, bodde mindre enn 10 mil unna eller mer enn 10 mil unna byene. Sammenhengene mellom barnas utdanning og familiekarakteristika vises både i tabell og grafisk.

Tabell 3.7 Andel i gruppen med høyere utdanning etter ulike familiekarakteristika.

Andel med høyere utdanning når	Menn		Kvinner	
	Obs	Gj.snitt	Obs	Gj.snitt
Ingen av foreldrene har høyere utdanning	3366	0,25	3266	0,29
Minst en av foreldrene har høyere utdanning	916	0,61	932	0,71
Familie er i den laveste inntektsgruppen	1460	0,2	1367	0,23
Familie er i den midterste inntektsgruppen	1406	0,29	1421	0,37
Familie er i den høyeste inntektsgruppen	1416	0,50	1410	0,56
Familie bor i Bergen / Trondheim	2133	0,36	2109	0,42
Familie bor mindre enn 10 mil unna Bergen / Trondheim	1415	0,30	1400	0,34
Familie bor mer enn 10 mil unna Bergen / Trondheim	735	0,30	689	0,35

¹⁷ Jamfør innledning.

Det ser ut til at det er sterk sammenheng mellom foreldrenes inntekt/utdanning og barnas deltagelse i høyere utdanning. I alle tilfeller har suksessgruppen¹⁸ de største andelene av individer med høyere utdanning. Spesielt ser sammenhengen mellom foreldres og barns utdanning ut til å være sterk. Gruppen som er bosatt i Bergen/Trondheim har også i større grad høyere utdanning enn de andre gruppene. Spesielt gjelder dette for kvinnene.¹⁹

Grafisk fremstilling av sammenhengen mellom familiekarakteristika og barnas utdanning.

At veksten i høyere utdanning er så mye sterkere for kvinner enn for menn, kan tyde på at det er andre mekanismer som styrer utdanningsforløpet hos kvinner. Det kan også tenkes at det er forskjell i sosial mobilitet mellom kjønnene. På grunn av dette skilles det mellom kjønn når sammenhengen mellom familiekarakteristika og barnas utdanning skal presenteres grafisk.

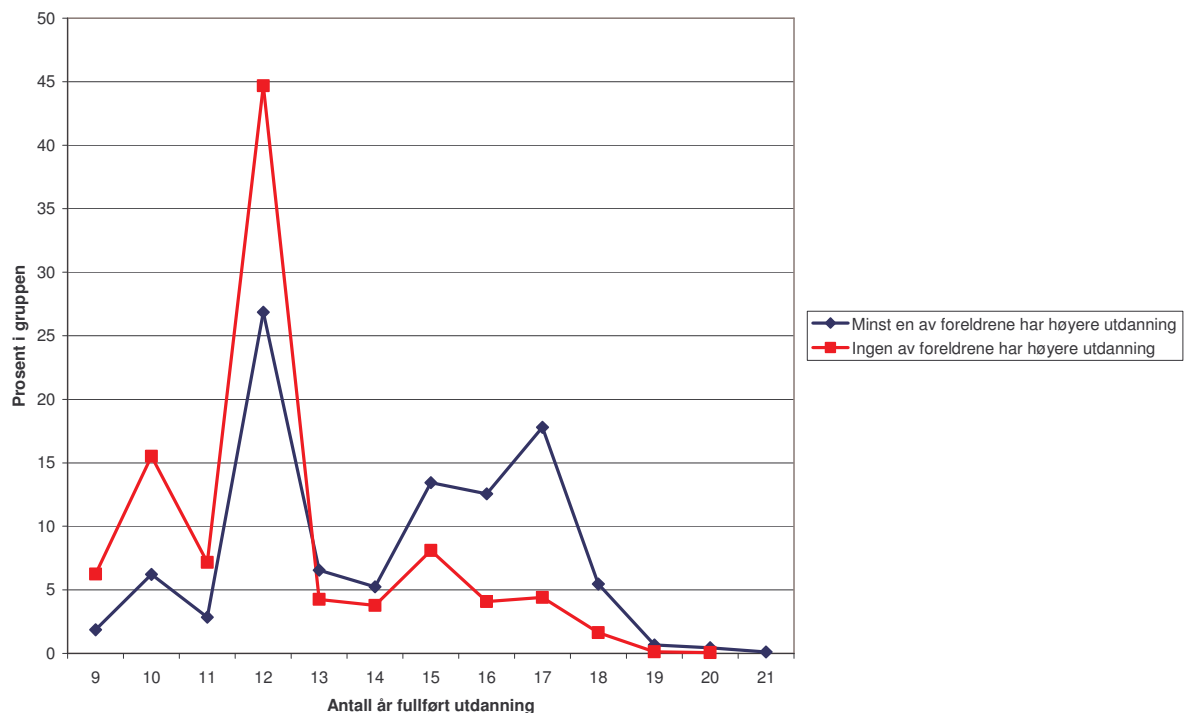
¹⁸ Med dette mener jeg gruppene man i følge tidligere forskning skulle anta hadde høyest utdanning.

¹⁹ Denne tendensen stemmer godt overens med funn Card (1993) og Frenette (2002 og 2003).

Foreldrenes utdanning og barns utdanning

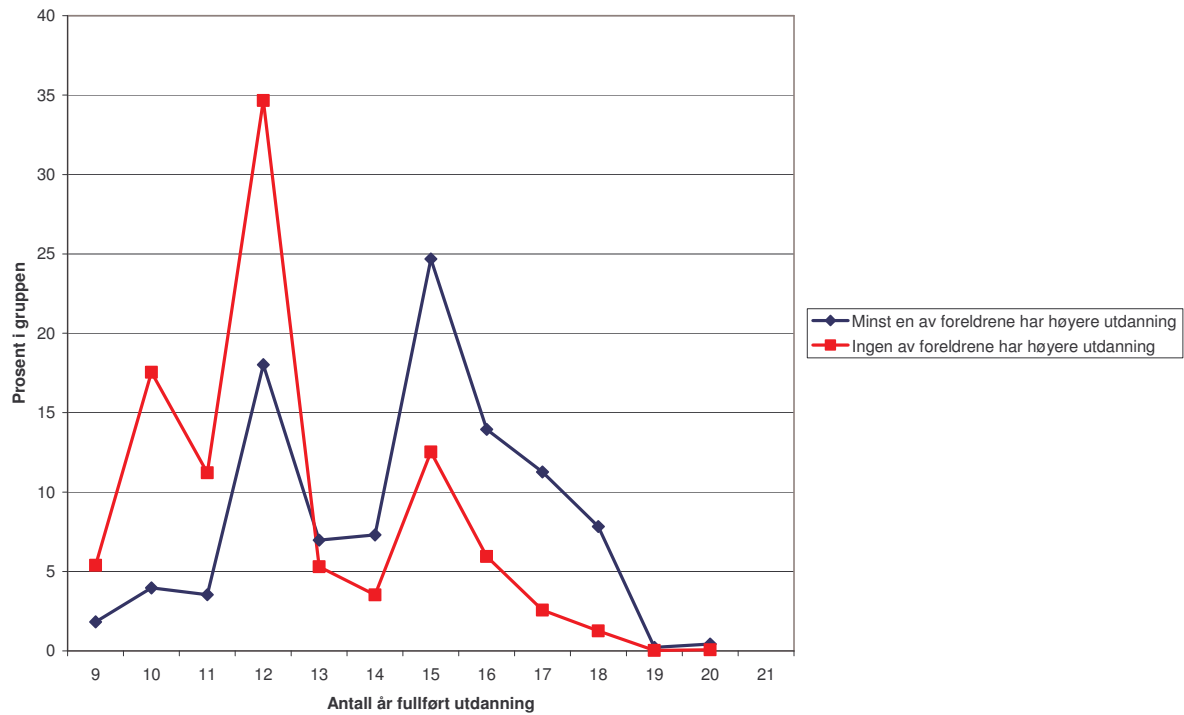
Jeg sammenligner her den høyeste fullførte utdanningen, målt i antall år, mellom barn hvor minst en av foreldrene har høyere utdanning mot barn hvor ingen av foreldrene har høyere utdanning. Punktene viser prosentvis andel i hver av gruppene som har det gitte antall år som høyeste fullførte utdanning.

Figur 3.2 Sammenheng mellom foreldrenes og barnas utdanning, menn.



Sammenhengen mellom foreldrenes utdanning og barnas utdanning ser ut til å være sterk. Barn av foreldre uten høyere utdanning slutter i større grad skolen etter 9-12 år, mens barn av foreldre med høyere utdanning i større grad fullfører høyere utdanning. Andelen med videregående skole som høyeste fullførte utdanning er 18 % større i gruppen hvor ingen av foreldrene har høyere utdanning. På masternivå er det også store forskjeller mellom gruppene. 13,5 % flere i gruppen hvor foreldrene har høyere utdanning enn i referansegruppen har så høy utdanning. Til tross for at gruppen med menn hvor foreldrene har høyere utdanning består av betydelig færre individer, 916 mot 3366, er antallet menn som har foreldre med høyere utdanning i flertall på masternivå. Dette er en indikasjon på at den sosiale reproduksjonen er sterk.

Figur 3.3 Sammenheng mellom foreldrenes og barnas utdanning, kvinner

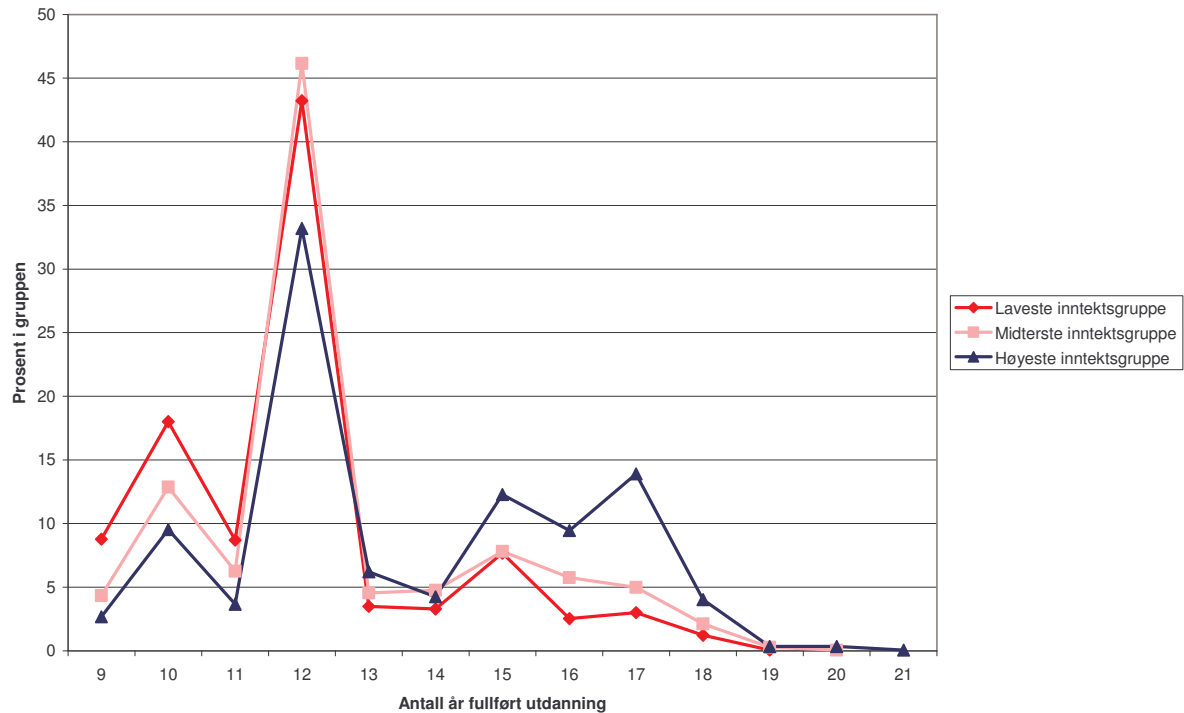


Utdanningsmønsteret for kvinner er annerledes enn for menn, men sammenhengene mellom barns utdanning og foreldreutdanning er likevel sterk. Betydelig flere av kvinnene med foreldre uten utdanning har kun 9-12 års utdanning, og betydelig flere av kvinnene hvor foreldrene har høyere utdanning tar selv høyere utdanning. Likevel viser figuren at kvinner fra begge gruppene i større grad enn menn tar høyere utdanning. Det er derfor vanskelig å si om det er det å være kvinne i seg selv som er positivt for deltagelsen til høyere utdanning, eller om kvinner fra lavt utdannede familier i større grad enn mennene klarer å løsrive seg fra sin sosioøkonomiske bakgrunn.

Familiens inntekt og barns utdanning

Jeg sammenligner her høyeste fullførte utdanning målt i antall år, for barn fra ulike inntektsgrupper. Utvalget er delt i tre etter familiens gjennomsnittelige inntekt.

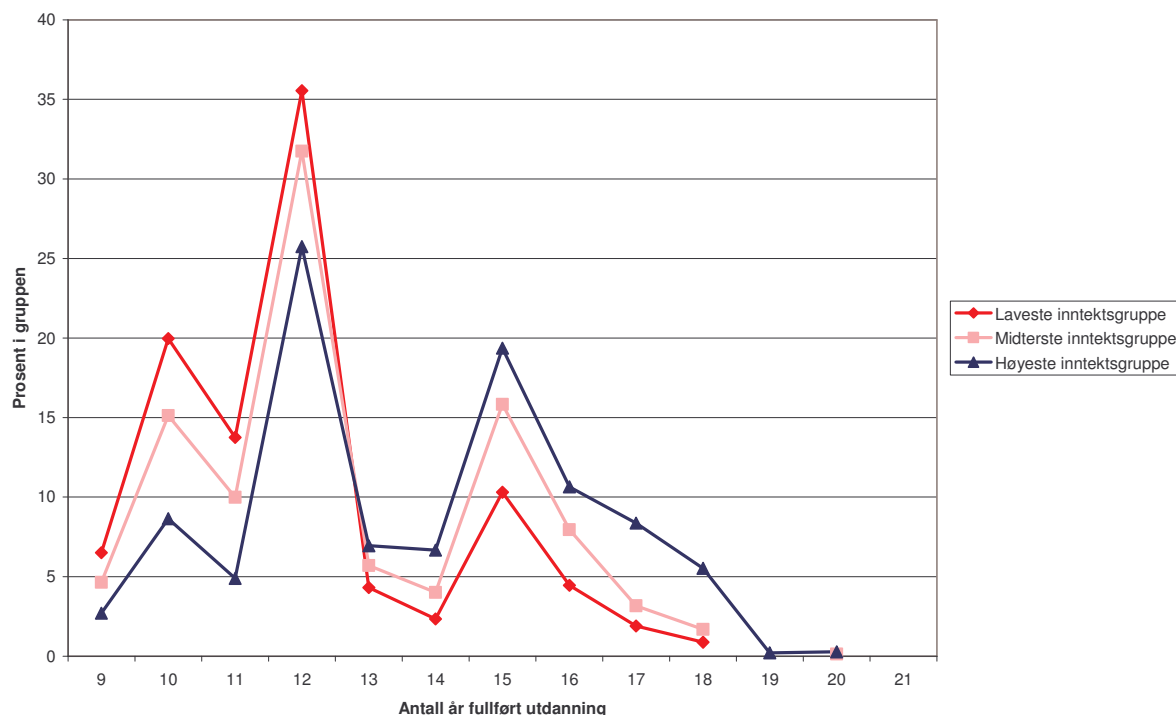
Figur 3.4 Sammenheng mellom barnas utdanning og familiens permanentinntekt, menn



Som i tilfellet med utdanning, ser vi en tydelig tendens i sammenheng mellom inntektsgruppene og barnas utdanning. Forskjellene mellom gruppene er likevel mindre. Dette kan skyldes at skillet mellom foreldre med og uten høyere utdanning utgjør et mer markant kulturell forskjell, enn ulikhet i inntekt.

Høy inntekt ser ut til å være sammenfallende med større sannsynlighet for å ta høyere utdanning. Barna i lavinntektsgruppen slutter i større grad av utdanningen sin etter obligatorisk skolegang, mens barn fra den midterste inntektsgruppen plasserer seg i midten også når det gjelder utdanning. Sammenligner man individene fra lavinntektsgruppen med de fra høyinntektsgruppen ser vi at 10 % flere fra den laveste inntektsgruppen slutter skolen etter videregående sammenlignet med høyinntektsbarna. På masternivå har situasjonen snudd og 10 % flere av de høyinntektsmennene har denne utdanningen sammenlignet med menn fra den laveste inntektsgruppen.

Figur 3.5 Sammenheng mellom barnas utdanning og familiens permanentinntekt, kvinner

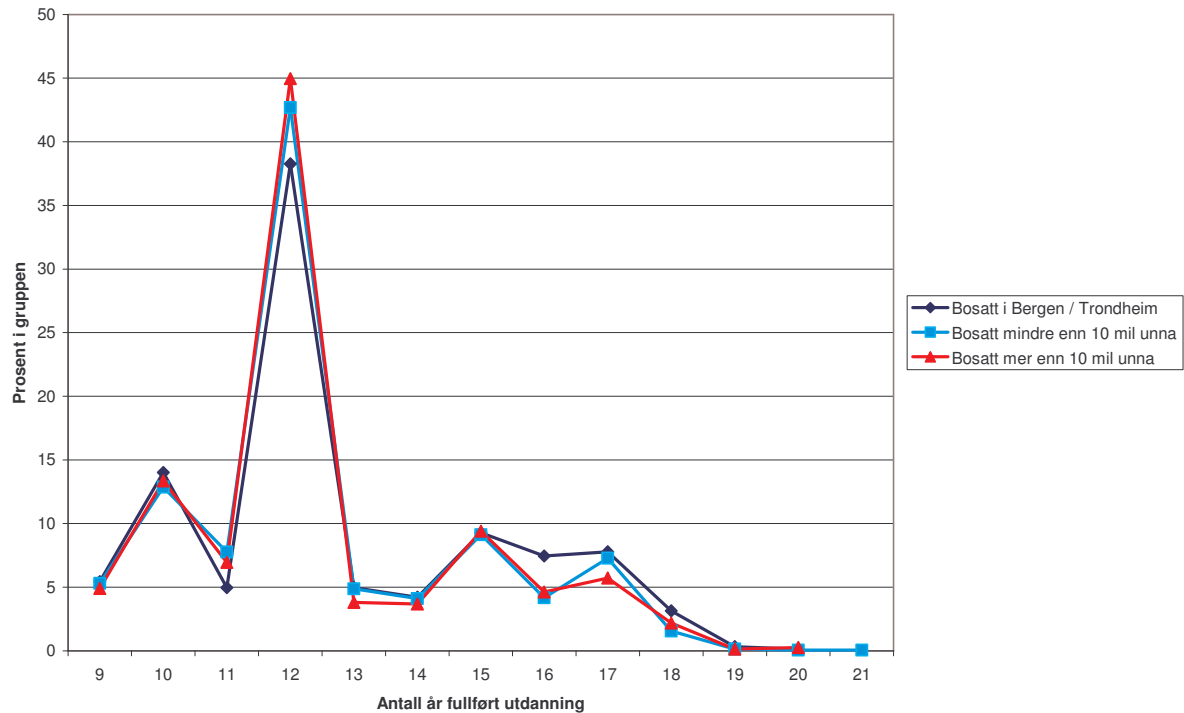


Høy inntekt forbindes også med høyere utdanning for kvinner. Det ser likevel ut til at det er mindre forskjell mellom høyinntektsbarn og de fra den midterste inntektsgruppen hos kvinnene sammenlignet med mennene. Etter 12 års utdanning har 9 % flere av lavinntektskvinnene sluttet skolen enn hva man finner i høyinntektsgruppen. Tilsvarende har 10 % flere av kvinnene i familier med høy inntekt bachelorgrad, sammenlignet med gruppen hvor foreldrene har lav inntekt. Nesten ingen av kvinnene fra de to laveste innteksgruppene har høyere utdanning enn 18 år. Kvinnene fra den høyeste inntektsgruppen er også i større grad enn mennene fra den laveste- og midterste inntektsgruppen representert på de høyeste utdanningene.

Avstand og barnas utdanning

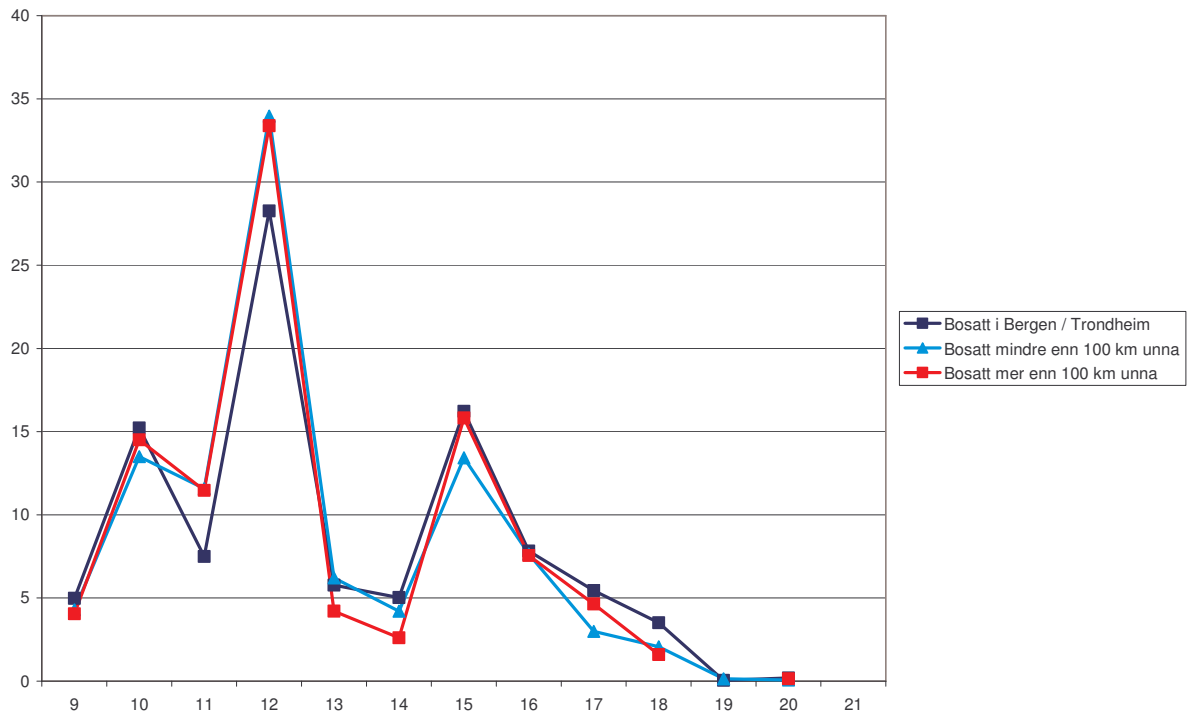
Her sammenlignes høyeste antall fullførte år med utdanning for menn og kvinner fra Bergen/Trondheim med individer bosatt mindre enn 10 mil unna, og mer enn 10 mil unna.

Figur 3.6 Sammenheng mellom avstand til utdanningsentrum og barnas utdanning, menn



Avstand ser ut til å ha negativ effekt på utdanningslengden. Lang avstand til høyere utdanningsinstitusjoner er sammenfallende med lavere utdanning. Studere man mennene som har videregående skole som høyeste fullførte utdanning, er forskjellen mellom bybarna og gruppen bosatt lengst vekke nesten 7 %. På bachelornivå er andelen sammenfallende for alle gruppene, men etter 16 års utdanning er andelen menn fra Bergen/Trondheim sterkere representert. Forskjellen er her nesten 3 %.

Figur 3.7 Sammenheng mellom avstand til utdanningsentrum og barnas utdanning, kvinner

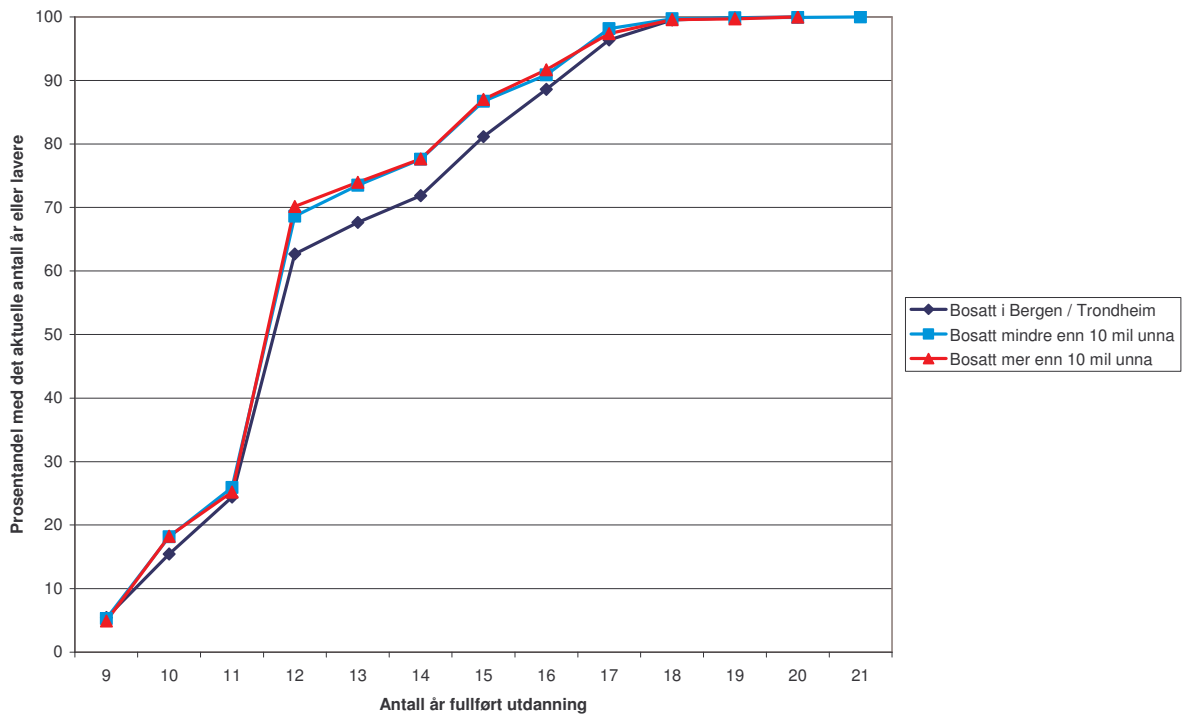


Avstand mellom hjemsted og Bergen/Trondheim ser ut til å ha litt andre effekter på utdanningen til kvinner enn for menn. Kvinnene bosatt i byene har fortsatt jevnt over høyere utdanning enn de andre. Flere av kvinnene bosatt langt borte slutter etter videregående skole. Skillet mellom gruppene er da 5 %. Kvinnene bosatt lengst vekk, kommer likevel sterkt tilbake, og ved de høyere utdanningene er andelen fra denne gruppen nesten like høy som hos kvinnene oppvokst i byene. Gruppen bosatt mindre enn 10 mil unna, ser ut til å være utdanningstaperne.

Kumulativ fordeling mellom avstand og antall år med utdanning

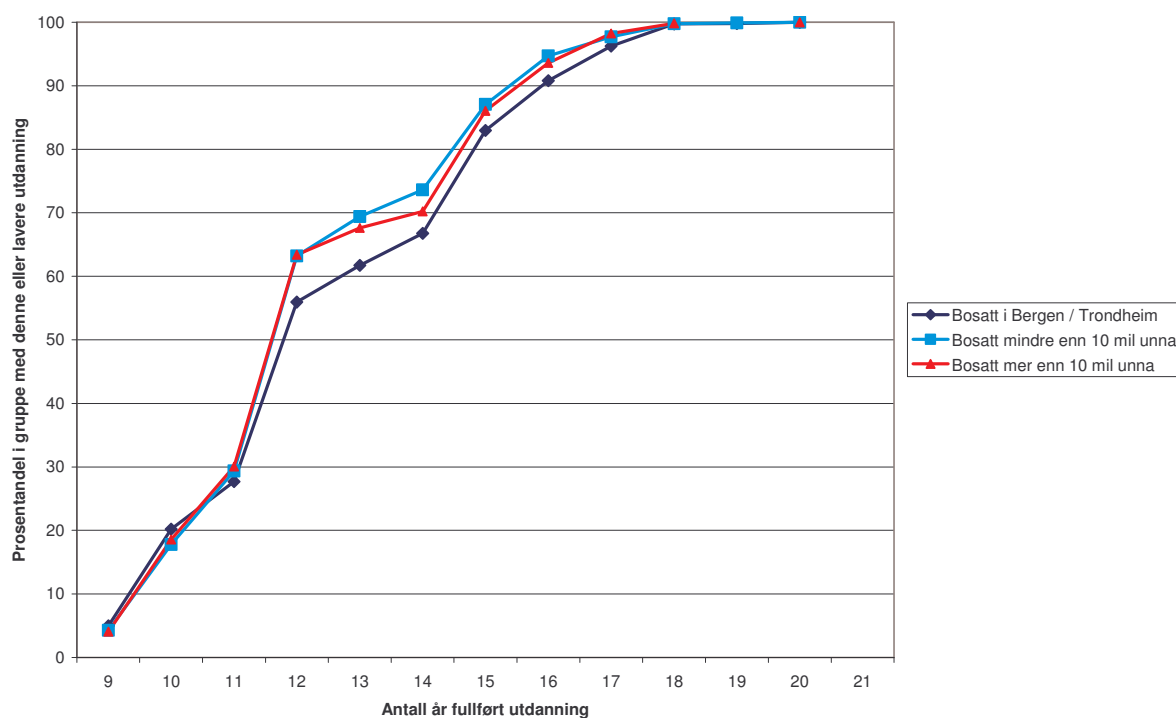
Figuren viser den kumulative fordelingen av antall år med utdanning for de ulike avstandsgruppene.

Figur 3.8 Kumulativ fordeling av antall år høyeste fullførte utdanning og avstand, menn.



Andel som har maksimalt 12 års utdanning er nesten 8 % lavere for menn bosatt i Bergen/Trondheim, enn for mennene bosatt mer enn 10 mil unna. Grafen viser at flere av de bosatt i Bergen / Trondheim har høyere utdanning. Forskjellene mellom gruppene bosatt utenfor byene ser ut til å være små. Denne fordelingen kan tyde på at det den sosiale effekten av å bo et stykke unna høyere utdanningsinstitusjoner, og ikke avstanden i seg selv som gjør at folk bosatt lengre unna tar mindre utdanning.

Figur 3.9 Kumulativ fordeling av antall år høyeste fullførte utdanning og avstand, kvinner



Andel kvinner med maksimalt 12 års utdanning er også for kvinnene ca 8 % lavere enn for mennene. Forskjellene mellom gruppene er relativt stor frem til 15 års utdanning. Deretter jevner det seg litt ut, men kvinnene fra Bergen/Trondheim har hele tiden mest utdanning. Forskjellene mellom gruppene bosatt utenfor byene er ikke veldig stor, men det ser ut til at de bosatt lengst vekk har litt høyere utdanning enn de bosatt i midten. Dette kommer spesielt til syne ved 14 års utdanning.

Oppsummering av sammenhenger mellom barns utdanning og familiekarakteristika

Tabellene og figurene i den deskriptive delen tyder på sentrale konsepter som ble introdusert i litteraturkapittelet, stemmer overens med virkeligheten for denne kohorten. I alle tilfeller indikerer høy foreldreutdanning/familieinntekt og lav avstand til nærmeste høyere utdanningsinstitusjon at barna har flere antall år med utdanning enn referansegruppene.

KAPITTEL 4: EMPIRISKE LIGNINGER

De empiriske analysene baserer seg på tre ulike regresjonsmodeller. Selv om modellen er spesifisert, er det fortsatt en del usikkerhetsmomenter knyttet til hvilke begrensninger dette særskilte datamaterialet legger. I del to av kapitlet testes derfor modellene for heteroskedastisitet, chow-test og korrelasjonene mellom variablene sjekkes, før regresjonsmodellene som faktisk estimeres presenteres.

4.1 Regresjonsmodeller

Det empiriske arbeidet består av tre modeller. Modell 1 benyttes til å etablere generelle sammenhenger mellom barnas utdanning og familiekarakteristika. Inntekt er her oppgitt som familiens gjennomsnittelige permanente inntekt, og avstand er ikke inkludert. I modell to ønsker jeg å undersøke i hvilken grad inntekt i ulike perioder har forskjellig betydning for evneutviklingen. Inntekt splittes derfor opp i mindre inntektsvinduer. Avstand er heller ikke her inkludert. Modell tre skal fange opp hvilken effekt avstand har på utdanningsdeltagelsen. Avstandsvariabelen er derfor inkludert i regresjonen. Inntekt er her permanentinntekt.

4.1.1 Lineære sannsynlighetsmodeller

Avstandsmatrisen baserer seg på distanse mellom hjemsted og nærmeste høyere utdanningsinstitusjon. Skal man evaluere effekten avstand har på høyere utdanning må måten utdanning måles på, korrespondere med måten avstandsmatrisen er utformet på. Utdanningsvariabelen i regresjonen er derfor en binær variabel som indikerer om individet har høyere utdanning eller ikke.

Når den avhengige variabelen er binær, er modellen en sannsynlighetsmodell.²⁰ Slike modeller kan estimeres på forskjellige måter, blant annet som lineære sannsynlighetsmodeller (LPM). Koeffisientene vil her rapportere hvordan endringer i forklaringsvariablene påvirker sannsynligheten for at individet har høyere utdanning. Problemet med denne måten å estimere på, er at den tilpassede sannsynligheten kan være mer eller mindre enn null og en, noe som strider med sannsynlighetsbegrepet. Den partielle effekten av forklaringsvariabelen som rapporteres er også konstant. Dette er ikke nødvendigvis tilfelle i virkeligheten.

²⁰ Se appendiks A.

Disse vanskelighetene kan omgås ved å bruke kumulative distribusjonsfunksjoner, for eksempel probitmodeller,²¹ til å estimere regresjonslikningene. Svakheten med denne type modeller er at koeffisientene er vanskelig å tolke.

I denne oppgaven benytter jeg en kombinasjon av de to. Regresjonsmodellene estimeres lineært, men probitmodellen brukes likevel som en forsikring for at den lineære metoden fungerer. I den grad de marginale resultatene fra probitmodellen²² ikke avviker mye fra koeffisientene i den lineære modellen, tyder dette på at forenklingen det er å estimere sannsynlighetsmodeller lineært kan rettferdiggjøres.

Modell 1 - Generelle sammehenger

$$Y_i = \beta_0 + \beta_1 I + \beta_2 U_m + \beta_3 U_f + \beta_4 S + \beta_5 A_m + \beta_6 A_f + \varepsilon$$

Notasjon

Y_i = Dummyvariabel som beskriver om individet har høyere utdanning eller ikke²³.

I = Familieinntekt mens barnet er 0-19 år.

U_m = Mors utdanningsnivå

U_f = Fars utdanningsnivå

S = Antall barn i familien

A_m = Mors alder

A_f = Fars alder

ε = Feilledd

Modell 2 - Effekt av inntekt i ulike perioder

$$Y_i = \beta_0 + \beta_1 I_{0-19} + \beta_2 I_{0-6} + \beta_3 I_{7-12} + \beta_4 I_{13-15} + \beta_5 I_{16-19} + \beta_6 U_m + \beta_7 U_f + \beta_8 S + \beta_9 A_m + \beta_{10} A_f + \varepsilon$$

Notasjon

²¹ Probitmodeller er en modelltype som baserer seg på normale kumulative distribusjonsfunksjoner. Der lineære modeller angir en rett linje gjennom punktene, og derfor kan gi predikerte verdier over og under 0 og 1, vil en probit-modell "skvise" forholdet mellom uavhengige og avhengige variabler innefor et spekter av [0,1]. Siden funksjonen ikke er lineær, er heller ikke $dF(x)/dx$ konstant. Dette gjør at de partielle effektene av forklaringsvariablene kan ha forskjellig verdi.

²² Funksjonen dprobit i STATA rapporterer endringen i sannsynligheten for en liten endring i hver uavhengige kontinuerlige variabel.

²³ Dummy har verdi = 1 hvis individet har høyere utdanning, 0 hvis ikke.

I_{0-19} = Familieinntekt mens barnet er 0-19 år²⁴

I_{0-6} = Familieinntekt mens barnet er 0-6 år

I_{7-12} = Familieinntekt mens barnet er 7-12 år

I_{13-15} = Familieinntekt mens barnet er 13-15 år.

I_{16-19} = Familieinntekt mens barnet er 16-19 år

Modell 3 - Effekt av avstand på deltagelse i høyere utdanning

$$Y_i = \beta_0 + \beta_1 I + \beta_2 U_m + \beta_3 U_f + \beta_4 S + \beta_5 A_m + \beta_6 A_f + \beta_7 D + \varepsilon$$

Notasjon

D = Avstand i km fra hjemsted til nærmeste utdanningsentrum

4.2 Tester av regresjonene

4.2.1 Multikollinearitet

Multikollinearitet vil si at det er høy korrelasjon mellom de uavhengige variablene. Effektene av dette er at man får ustabile koeffisienter, og at tolkningen av variablene kan bli vanskeligere. Har man høy korrelasjon mellom to uavhengige variabler, er det vanskelig å estimere den partielle effekten hver forklaringsvariabel har på den avhengige variabelen, siden disse blandes inn i hverandre. Multikollinearitet kan også gi store endringer i koeffisientverdiene ettersom variabler inkluderes/ekskluderes fra regresjonene. Det er ikke satt en bestemt grense for hvor høy korrelasjon det kan være mellom variablene før det defineres som multikollinearitet, og dette vurderes fra situasjon til situasjon. For å unngå dette problemet sjekkes korrelasjonene mellom variablene før modellen estimeres.

Tabell 4.1 Korrelasjoner mellom variablene som benyttes i modell 1 og 3.

Variabel	Høyere utdanning	Antall barn	Mors alder	Fars alder	Mors utd.nivå	Fars utd.nivå	Inntekt	Avstand
Høyere utdanning	1,00							
Antall barn	-0,09	1,00						
Mors alder	0,06	0,51	1,00					
Fars alder	0,03	0,48	0,82	1,00				
Mors utdanningsnivå	0,33	-0,14	0,02	-0,02	1,00			
Fars utdanningsnivå	0,36	-0,12	0,03	-0,04	0,48	1,00		
Inntekt	0,28	-0,10	0,00	-0,08	0,40	0,48	1,00	
Avstand	-0,06	0,05	-0,08	-0,01	-0,12	-0,15	-0,20	1,00

²⁴ Variablene er diskonterte logaritmiske gjennomsnittsinntekter.

Korrelasjonene holder seg i hovedsak under 0,5, som er et akseptabelt nivå. Sammenhengen mellom mors og fars alder er derimot 0,82. Dette er svært høyt. Tidligere forskning har lagt lite vekt på fars alder når barns utdanningsdeltakelse skal forklares, og for å unngå multikollinearitet dropper jeg derfor denne variabelen fra ligningene. Effekten av foreldrenes alder måles nå gjennom variabelen for mors alder.

I modell 2 estimeres effekten inntekt har på sannsynligheten for deltagelse til høyere utdanning, i ulike perioder av barnets oppvekst. Det er naturlig å tenke seg at det er sammenheng mellom familiens inntekt i ulike perioder siden høy lønn i en periode ofte etterfølges av høy lønn neste periode. Et problem med å dele opp inntektene i mindre vinduer kan være at man får høye korrelasjoner mellom variablene. Av korrelasjonstabellen ser man at dette er tilfelle.

Tabell 4.2 Korrelasjon mellom inntektsvariablene

Variabel	(0-19)	(0-6)	(7-12)	(13-15)	(16-19)
Inntekt (0-19)	1,00				
Inntekt (0-6)	0,73	1,00			
Inntekt (7-12)	0,89	0,63	1,00		
Inntekt (13-15)	0,84	0,48	0,76	1,00	
Inntekt (16-19)	0,67	0,28	0,45	0,53	1,00

For å minimere samvariasjonproblemer utelukkes permanentinntekt og inntekt mens barnet går på ungdomsskolen når modellen skal estimeres. Siden modellen ikke inneholder alle inntektsvinduerne er den underspesifisert. Koeffisientene er derfor ikke nøyaktige, men resultatene kan likevel benyttes til å si noe om i hvilke perioder inntekt er viktig for utdanningsdeltagelse.

4.2.2 Chow-test

Kjønn er et viktig element i analysen. Hvordan jeg forholder meg til denne variabelen avhenger om menn og kvinner har like terskelverdier. Er verdiene like, inkluderer jeg kjønn som en dummy. Avviker verdiene fra hverandre må det spesifiseres to modeller, en for menn og en for kvinner. Jeg anvender Chow-test til å undersøke dette. Testens nullhypotese er at terskelverdiene for menn og kvinner er identiske.

Formelt blir dette, hvor m indikerer mann, og k kvinne:

$$\beta_{0,m} = \beta_{0,k}$$

$$\beta_{1,m} = \beta_{1,k}$$

$$\beta_{2,m} = \beta_{2,k}$$

$$H_0: \beta_{3,m} = \beta_{3,k}$$

$$\beta_{4,m} = \beta_{4,k}$$

$$\beta_{5,m} = \beta_{5,k}$$

$$\beta_{6,m} = \beta_{6,k}$$

H1: Minst en av verdiene er forskjellig.

Testobservatoren finner man ved følgende formel

$$F = \left[\frac{SSR_R - (SSR_K + SSR_M)}{SSR_K + SSR_M} \right] * \frac{[n - 2(k + 1)]}{k + 1}$$

SSR_R er summen av kvadrerte residualer for regresjonen som ikke tar hensyn til kjønn

SSR_K er summen av kvadrerte residualer for kvinneregresjonen

SSR_M er summen av kvadrerte residualer for manneregresjonen

n er antall frihetsgrader

$k+1$ er antall restriksjoner

Innsatt i formel får man:

$$F = \frac{[1603,63 - (810,71 + 783,57)] * \frac{8466}{7}}{810,71 + 783,57} = 7,09$$

Fra F-tabell finner vi kritisk verdi ved 5 % signifikansnivå $F_{6,\infty} = 2,10$. Her er testobservatoren 7,09 - H_0 forkastes. Terskelverdiene er forskjellige, og det må spesifiseres en modell for kvinner og en for menn.

4.2.3 Heteroskedastisitet

Heteroskedastisitet er systematisk variasjon i størrelsen til residualene etter verdien av variablene. Implikasjoner av dette er at standardfeilen til regresjonskoeffisientene blir skjevt estimert, og t og F verdier ikke er korrekte. Dette innebærer at man ikke kan lage konfidensintervaller eller hypoteseteste. Disse problemene kan omgås ved å bruke robuste standardfeil. Heteroskedastisitet er også en indikasjon på at modellen kan estimeres mer effektivt, men skal man forbedre dette må man kjenne heteroskedastisiteten.

I den grad lineære sannsynlighetsmodeller har forklaringskraft vil det alltid eksistere heteroskedastisitet.²⁵ Benytter likevel Breusch -Pagan test for å sjekke for heteroskedastisitet i min modell. Nullhypotesen i denne testen er at variansen til residualene er homogen.

Formelt blir dette:

$$H_0: \text{Var}(\varepsilon | I_{0-19}, U_m, U_f, A_m, A_f, D) = \sigma^2$$

$$H_1: \text{Var}(\varepsilon | I_{0-19}, U_m, U_f, A_m, A_f, D) \neq \sigma^2$$

Ved lave p-verdier må vi forkaste nullhypotesen og akseptere alternativhypotesen om at variansen ikke er homogen.

Resultat fra STATA:

$$\text{Chi2}(1) = 67,89$$

$$\text{Prob} > \text{chi2} = 0,0000$$

Resultatene fra STATA gir p-verdier lik null, og H_0 må forkastes. Regresjonen har heteroskedastisitet. For å kompensere for dette benytter jeg robuste standardfeil når modellene estimeres.

4.3 Estimerte modeller

Korrelasjonstabellene og testene som ble kjørt på modellene, viste at modellene som skal estimeres må gjennom enkelte justeringer. Fars alder utelukkes fra alle modellene. Permanentinntekt og familiens inntekt mens barna går på ungdomsskolen utelukkes fra modell to. Det estimeres en modell for kvinner og en for menn, og det benyttes robuste standardfeil.

De empiriske ligningene som estimeres er derfor:

Modell 1

$$Y_{i,j} = \beta_0 + \beta_1 I + \beta_2 U_m + \beta_3 U_f + \beta_4 S + \beta_5 A_m + \varepsilon$$

²⁵ Dette vises i appendiks A.

Modell 2

$$Y_{i,j} = \beta_0 + \beta_1 I_{0-6} + \beta_2 I_{7-12} + \beta_3 I_{16-19} + \beta_4 U_m + \beta_5 U_f + \beta_6 S + \beta_9 A_m + \varepsilon$$

Modell 3

$$Y_{i,j} = \beta_0 + \beta_1 D + \beta_2 I_{0-19} + \beta_3 U_m + \beta_4 U_f + \beta_5 S + \beta_6 A_m + \varepsilon$$

Notasjon som før,

j = mann eller kvinne,

KAPITTEL 5: RESULTATER

I dette kapittelet presenterer jeg resultatene fra de estimerte regresjonene, og funnene vurderes opp mot tidligere forskning. Kapittelet inneholder også en analysedel med diskusjon om samvariasjonene skal tolkes som en konsekvens av kausalitet eller seleksjon. Til sist evaluerer jeg den sosiale utdanningsmobiliteten mellom barn og foreldre.

5.1 Regresjonsresultater

De lineære sannsynlighetsmodellene estimeres ved hjelp av minste kvadraters metode (MKM).²⁶ Det predikerte utfallet er sannsynligheten for at individet har høyere utdanning. For å sjekke validiteten til resultatene fra den lineære sannsynlighetsmodellen, har jeg også inkludert de marginale resultatene fra probitmodellen.²⁷

5.1.1 Modell 1

Alle regresjonene er basert på 4282 observasjoner av menn og 4198 observasjoner av kvinner.

Tabell 5.1 Resultater fra modell 1. Sammenheng mellom høyere utdanning og familiekarakteristika. Lineær estimering.

Høyere utdanning Variabel	Menn		Kvinner	
	Koeffisient	St.avvik	Koeffisient	St.avvik
Permanentinntekt	0,1331	0,0214	0,1340	0,0238
Mors utdanningsnivå	0,0846	0,0087	0,0969	0,0085
Fars utdanningsnivå	0,0723	0,0064	0,0801	0,0065
Antall barn i familien	-0,0319	0,0071	-0,0366	0,0074
Mors alder	0,0088	0,0014	0,0078	0,0015

Tabell 5.2 Resultater fra modell 1. Sammenheng mellom høyere utdanning og familiekarakteristika. Marginaleffekter fra probitmodell.

Høyere utdanning Variabel	Menn		Kvinner	
	Koeffisient	St.avvik	Koeffisient	St.avvik
Permanentinntekt	0,1633	0,0296	0,1667	0,0337
Mors utdanningsnivå	0,0809	0,0088	0,1043	0,0097
Fars utdanningsnivå	0,0714	0,0067	0,0850	0,0072
Antall barn i familien	-0,0378	0,0083	-0,0434	0,0090
Mors alder	0,0098	0,0016	0,0089	0,0017

²⁶ Metoden går ut på å minimere summen av kvadrerte residualer.

²⁷ Resultatene fra den opprinnelige probit-modellen finnes i appendiks B.

Alle resultatene er skarpt bestemt, og regresjonene har forklaringskraft på mellom 16 og 19 prosent²⁸. Fortegnene på koeffisientene er som forventet i forhold til tidligere studier. Resultatene fra den estimerte modellen tyder på foreldrenes inntekt og utdanning er de viktigste kjennetegnene for deltagelse i høyere utdanning hos barna.

Den lineære sannsynlighetsmodellen ser ut til å stemme relativt godt overens med probitmodellen. Den marginale effekten av inntekt på sannsynligheten for deltagelse i høyere utdanning er likevel noe høyere i probitmodellen. Dette kan tyde på at modell 1 ikke er helt lineær. For gjennomsnittsindividet ser det derfor ut til at effekten av inntekt blir underestimert i den lineære modellen. Ut i fra resultatene som er presentert her er det vanskelig å si noe om hvordan denne effekten er for andre grupper.²⁹

Begge kjønn har inntektskoeffisient på ca 0,13. Siden inntekten er oppgitt logaritmisk betyr dette at en endring i permanentinntekt på 1 % øker sannsynligheten for at individet har høyere utdanning med 0,001331 prosentpoeng. For gjennomsnittsfamilien er en prosent av inntekten 22 000 kroner³⁰. Studerer man inntektsfordelingene mellom familiene ser vi at 80 % av familiene har deflaterte inntekter mellom 150 000 og 300 000 kroner³¹. Siden forskjellene i inntekt mellom familien er liten, er det lite trolig at inntekt i stor grad kan benyttes til å forklare forskjeller i sannsynlighet for deltakelse i høyere utdanning for dette utvalget.³²

Foreldrenes utdanning ser ut til å ha sterk effekt på utdanningsforløpet til barna. På grunn av den tette inntektsfordelingen mellom familiene, er foreldrenes utdanning den variabelen som i størst grad kan forklare forskjeller i utdanning hos utvalget.³³

Resultatene tyder på at foreldrenes utdanning er noe viktigere for kvinner enn for menn, og at mors utdanning er noe viktigere enn fars, for begge kjønn. Et ekstra utdanningsnivå til mor,

²⁸Forklaringskraften til sannsynlighetsmodeller er generelt lav. Dette henger sammen med at den binære variabelen kun kan ha verdiene null og en, mens det predikerte utfallet kan ha alle verdier. Regresjonslinjen vil derfor sjelden passe svært bra til de aktuelle dataene.

²⁹ Disse effektene kan man finne ved å sette inn resultatene fra probitmodellen i regresjonsligningen, og regne ut for interessant verdier av forklaringsvariablene.

³⁰ 1998-kroner.

³¹ Fordelingen vises i appendiks C.

³² Raaum (2003) peker på at 100 000 kroner ekstra i årlig familieinntekt utgjør et ekstra skoleår for barnet. Effekten av inntekt reduseres når man regner med foreldrenes utdanning. Deltagelse i høyere utdanning er et mindre nyansert mål enn en måling av skoleår. Det kan derfor tenkes at effekten av inntekt på sannsynligheten for deltagelse i høyere utdanning ikke fullt ut viser effekten inntekt har på antall år med utdanning.

³³ Dette stemmer overens med Raaum (2003).

øker sannsynligheten for at barnet tar høyere utdanning med 8,4 prosentpoeng for menn, og 9,6 prosentpoeng for kvinner. Har mor treårig høyskole- eller universitetsutdanning vil dette øke sannsynligheten for at barnet har høyere utdanning med 26-30 prosentpoeng sammenlignet med individer hvor mor kun har obligatorisk skolegang. Fars utdanning har nesten like sterk effekt. Et ekstra utdanningsnivå øker sannsynligheten for at sønnene har høyere utdanning med 7,2 prosentpoeng, og 8 prosentpoeng for døtrene.

Søsken virker negativt på sannsynlighet for at man har høyere utdanning.³⁴ Et ekstra barn reduserer sannsynligheten med i overkant av 3 prosentpoeng. Effekten av å ha tre søsken vil derfor være en reduksjon i sannsynligheten med 10 prosentpoeng, noe som omtrent tilsvarer et utdanningsnivå for mor.

Mors alder har signifikant effekt på sannsynlighet for utdanningsdeltakelse. Effekten per år er liten, men for to ellers like individer vil en ti år eldre mor øke sannsynligheten for å ha høyere utdanning med 7-8 prosentpoeng.

For å sjekke representativiteten til utvalget ble regresjonen også kjørt for hele kohorten.³⁵ Ser man bort fra forskjeller i betydningen av inntekt for sannsynligheten for kvinners deltagelse i høyere utdanning, avviker ikke resultatene fra denne regresjonen nevneverdig. Resultatene for utvalget er derfor representativ for hele kohorten.

³⁴ Raaum (2003) fant at søskeneffekten ikke er monoton, å ha ett søsken øker utdanningslengden. At søsken kan ha forskjellig effekt på sannsynligheten for høyere utdanning er det ikke tatt høyde for her.

³⁵ Disse resultatene finnes i appendiks B.

5.1.2 Modell 2

Tabellen viser sammenhengen mellom barnas utdanning og variablene når inntekten er delt opp i mindre inntektsvinduer.

Tabell 5.3 Modell 2. Sammenheng mellom høyere utdanning og ulike inntektsvariabler

Høyere utdanning	Menn		Kvinner	
	Koeffisient	St.avvik	Koeffisient	St.avvik
Inntekt (0-6)	0,0592	0,0305	0,0510	0,0244
Inntekt (7-12)	0,0244	0,0195	0,0055	0,0241
Inntekt(16-19)	0,0362	0,0107	0,0637	0,0107
Mors utdanningsnivå	0,086	0,0086	0,0956	0,0084
Fars utdanningsnivå	0,0738	0,0065	0,0819	0,0064
Antall barn i familien	-0,0336	0,007	-0,0362	0,0075
Mors alder	0,0088	0,0015	0,0081	0,0015

Forklaringskraften til modellen ligger mellom 16-18 prosent, og fortegnene stemmer overens med det som var forventet. Koeffisientene for inntekt før barnet har begynt på skolen er statistisk signifikant på 5 %-nivå for begge kjønn. Mennenes koeffisient for inntekt mens barnet går på videregående er statistisk signifikant på 1 %-nivå. Inntektskoeffisientene mens barna går på barneskolen er ikke statistisk signifikant.

Siden flere av inntektsvariablene allerede er utelatt, og barneskolevariabelen ikke er statistisk signifikant, er materialet som skal brukes til å analysere forskjeller i effekten av inntekt i ulike tidsperioder tynt. For likevel å få en indikasjon på om inntekt kan ha forskjellig betydning for sannsynligheten for deltagelse i høyere utdanning i ulike perioder, er modellen estimert med ett og ett inntektsvindu. Verdien av koeffisientene er da mest sannsynlig overdrevet siden ingen av de andre inntektsvariablene er med.

Tabell 5.4 Effekt av inntekt på høyere utdanning når variablene er inkludert hver for seg.

Høyere utdanning	Menn		Kvinner	
	Koeffisient	St.avvik	Koeffisient	St.avvik
Inntekt (0-6)	0,0871	0,0279	0,0791	0,0200
Inntekt (7-12)	0,0631	0,0143	0,0676	0,0199
Inntekt (13-15)	0,0802	0,0132	0,0705	0,0149
Inntekt(16-19)	0,0468	0,0100	0,0709	0,0098

Alle resultatene i disse modellene er skarpt bestemt.

Både resultatene fra modell 2 og den provisoriske estimeringen tyder på at det finnes visse kjønnsforskjeller i hvordan inntekt påvirker sannsynligheten for deltagelse i høyere utdanning.

For menn ser det ut til at familiens inntekt før de begynner på barneskolen og mens de går på ungdomsskolen har størst betydning for utdanningsdeltagelsen. At høy inntekt i perioden hvor kognitive evner formes er viktig, stemmer godt overens med teori.³⁶ Den høye koeffisienten på ungdomsskolen kan ha sammenheng med at dette er tidspunktet individene bestemmer seg for om de skal ta yrkesutdanning eller en mer allmennfaglig rettet utdanning. Yrkesvalgene i Norge er tradisjonelle, og flere menn enn kvinner har yrkesutdanning.³⁷ Et resultat av dette kan være at flere menn enn kvinner tar avgjørende yrkesvalg på ungdomsskolen. Selv om effekten er liten, kan koeffisienten tyde på at familiens økonomi kan ha noe å si for dette valget.³⁸ Sammenlignet med de andre periodene har familiens inntekt mens sønnene går på videregående skole relativt lite å si.

For kvinnene er resultatene mer stabile, og forskjellen mellom periodene ikke særlig store. Familiens inntekt før kvinnene begynner på skolen er viktigst, men ikke mye viktigere enn når de går på videregående. Også for kvinner er barneskolen den perioden familiens inntekt er minst viktig. Hvis kvinner i større grad enn menn tar utdanningsbeslutninger på et senere tidspunkt, kan samme argumentasjon som ble brukt for viktigheten av familieinntekt på ungdomsskolen for menn også benyttes for kvinner på videregående skole.

Resultatene fra modell 1 viser at koeffisienten for permanentinntekt er høyere enn alle de provisoriske inntektskoeffisienter³⁹ i modell 2. Dette kan tolkes som at langsiktige inntektseffekter er viktigere for utvalget enn kortsiktige.⁴⁰

³⁶ Jamfør Heckman (2005)

³⁷ Dette underbygges av statistikk fra SSB.

³⁸ Black, Devereux og Salvanes (2005) fant at kredittbegrensninger ikke har hatt avgjørende betydning for skolegangen til svakerestilte grupper i Norge. Familiens økonomiske situasjon kan likevel ha effekt på preferansene til individet.

³⁹ Verdien av disse er også overestimert siden ingen av de andre inntektsvariablene er inkludert.

⁴⁰ Dette samsvarer med resultater i Dearden (2004).

5.1.3 Modell 3

Tabellen viser regresjonsresultatene når avstand inkluderes i regresjonen med permanentinntekt.

Tabell 5.5 Modell 3. Sammenheng mellom høyere utdanning og avstand.

Høyere utdanning Variabel	Menn		Kvinner	
	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	0,0035	0,0013	0,0023	0,0013
Permanentinntekt	0,1426	0,2179	0,0138	0,2388
Mors utdanningsnivå	0,0849	0,0087	0,097	0,0085
Fars utdanningsnivå	0,0731	0,0064	0,0809	0,0065
Antall barn i familien	-0,0337	0,0071	-0,0373	0,0075
Mors alder	0,0092	0,0014	0,008	0,0015

Alle variablene er skarpt bestemt bortsett fra avstand som er statistisk signifikant på 1 %-nivå for menn og på 10%-nivå for kvinner.

Effekten av avstand på utdanningsdeltagelsen er positiv.⁴¹ Det vil si at sannsynligheten for at både kvinner og menn tar høyere utdanning øker når kilometeravstanden til nærmeste høyere utdanningsinstitusjon øker. Dette resultatet er svært overraskende. Både Card (1993) og Frenette (2002) har entydig vist negativ sammenheng mellom utdanningsdeltagelse og avstand.

Både figurene i den deskriptive statistikken, og korrelasjonstabellen viste negativ samvariasjon mellom utdanningsdeltakelse og avstand. At koeffisienten nå er positiv kan være et tegn på at det eksisterer multikollinearitet. For å undersøke mulige feilkilder inkluderer jeg variabel for variabel og sjekker fortegn og signifikans.

Variabel for variabel

Avstand har negativ og signifikant effekt på sannsynligheten for deltagelse i høyere utdanning for både menn og kvinner så lenge ikke en av variablene mors utdanningsnivå, fars utdanningsnivå eller permanentinntekt inkluderes. Alle variablene utenom avstand, er skarpt bestemt i disse regresjonene.

⁴¹ Regresjonene ble også kjørt med reisetid som mål på avstand. Dette ga samme fortegn og verdiene er var fortsatt signifikante, som når avstand måles i distanse. Det ble også sjekket om den positive effekten av avstand endret seg når også individer med fylkestilhørighet i Sogn og Fjordane og Nord-Trøndelag ble inkludert. Avstandsvariablen var fortsatt positiv og statistisk signifikant. Resultatene fra dette finnes i appendiks B.

Tabell 5.6 a)

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	-0,0047	0,0014	-0,055	0,0014

Avstand er skarpt bestemt.

Tabell 5.6 b)

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	-0,0042	0,0014	-0,0053	0,0014
Antall barn	-0,0339	0,0064	-0,0435	0,0069

Avstandsvariabelen til menn er statistisk signifikant på 1- % nivå, ellers er alle variabler skarpt bestemt

Tabell 5.6 c)

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	-0,0293	0,0014	-0,0385	0,0014
Antall barn	-0,0679	0,0075	-0,762	0,0079
Mors alder 1970	0,1267	0,0015	0,0127	0,0016

Avstand statistisk signifikant på 5 % nivå for menn og på 1 % for kvinner

Når foreldrenes utdanning eller inntekt inkluderes blir avstandsvariabelen i hovedsak positiv, og ikke lenger statistisk signifikant.

Tabell 5.6 d)

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	0,0001	0,0013	-0,0010	0,0014
Antall barn	-0,0438	0,0072	-0,0462	0,0077
Mors alder 1970	0,0098	0,0014	0,0092	0,0015
Mors utdanningsnivå	0,1474	0,0073	0,165	0,007

Avstand er ikke statistisk signifikant verken for menn eller kvinner

Tabell 5.6 e)

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	0,0013	0,0013	0,0009	0,0014
Antall barn	-0,0428	0,0072	-0,0502	0,0076
Mors alder 1970	0,0101	0,0014	0,0094	0,0015
Fars utdanningsnivå	0,1179	0,0054	0,1283	0,0052

Avstand er ikke statistisk signifikant verken for menn eller kvinner

Tabell 5.6 f)

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	0,0020	0,0013	0,0003	0,0014
Antall barn	-0,0558	0,0073	-0,0624	0,0077
Mors alder 1970	0,0121	0,0015	0,0112	0,0015
Permanentinntekt	0,3602	0,0238	0,3855	0,0258

Avstand er ikke statistisk signifikant verken for menn eller kvinner.

Avstand er negativ og signifikant både i tilfellene der avstand er den eneste forklaringsvariabelen, og når søsken og mors alder inkluderes en etter en, men både koeffisientverdien og signifikansnivået reduseres ettersom flere variabler kommer til. De andre variablene er skarpt bestemte og har forventet fortegn. Når enten mors utdanning, fars utdanning eller familiens inntekt inkluderes, endrer effekten av avstand seg. Koeffisienten skifter da fortegn, og er heller ikke statistisk signifikant. Inkluderes derimot alle disse variablene i regresjonen samtidig, blir avstandskoeffisientene positiv og signifikant. Siden avstand i utgangspunktet hadde negativ og signifikant effekt på den avhengige variabelen, er fortegnsendringen et tegn på at vi har et multikollinearitetsproblem.

Ut i fra korrelasjonstabellen burde ikke dette være noe problem. Korrelasjonen mellom avstand og foreldrenes utdanning og inntekt er mellom -0,12 og -0,19. Multikollinearitetsproblemet forteller derfor om svakheten ved å bruke mors bostedskommune i 1960 som utgangspunkt for avstandsmatrisen. Det er nærliggende å tenke at denne variabelen i virkeligheten forteller om effekten avstand har på foreldrenes inntekt og utdanning, i stedet for på barnas utdanningsdeltagelse. Dette aspektet påvirker selvsagt barna, men gjennom mekanismene foreldreutdanning/inntekt og ikke avstand. Som et resultat av dette, er effekten av avstand mellom hjemsted og utdanningssted, på 68-kohortens utdanningsdeltagelse uviss.

5.2 Seleksjon versus kausalitet

Skal resultatene tolkes eller anvendes videre må det vurderes om samvariasjonene mellom uavhengige og avhengige variabler skyldes kausalitet, eller om de er et uttrykk for seleksjon.

Kausalitet vil si at det er en direkte sammenheng mellom årsak og virkning. Denne tilnærmingen innebærer at endring av en forklaringsvariabel er den direkte årsaken til at den avhengige variablene endres. Resultatene fra modellene viste at det er positiv samvariasjon mellom foreldrenes inntekt og sannsynligheten for at barna har høyere utdanning. Den kausale forklaringen på dette, er at det er inntekten i seg selv som fører til økt sannsynlighet. Med dette resonnementet vil man ved å øke foreldrenes inntekt automatisk øke sannsynligheten for at barnet tar høyere utdanning.

Den alternative måten å tolke sammenhengene på, er ved seleksjon. Man antar da at forklaringsvariablene indikerer foreldrenes egnethet. Variabelen inntekt kan tjene som et eksempel på å forklare dette. Antar man at lønn avspeiler dyktighet, vil foreldre med høy lønn, ha bedre evner enn foreldre med lav lønn. Den positive samvariasjonen mellom foreldres inntekt og barnas utdanning henger derfor sammen med at høy lønn indikerer gode foreldreevner. Disse blir overført til barna gjennom arv og miljø. Det er derfor de bakenforliggende evnene som inntekten avspeiler, og ikke lønnen i seg selv, som gjør at høy foreldreinntekt er forbundet med høy utdanning hos barna. Å gi en familie mer penger vil, med denne tilnærmingen, ikke øke sannsynligheten for at barna har høyere utdanning.

I hvilken grad sammenhengene mellom variablene skyldes kausalitet eller er et resultat av seleksjon, er på mange måter en forlengelse av spørsmålet om evneutvikling kan tilskrives arv eller miljø. Som i tilfellet med evner, kan det tenkes at sammenhengen mellom utdanningsdeltagelse og familiekarakteristika er et samspill mellom kausalitet og seleksjon. De ulike forklaringsvariablene kan også kreve forskjellig tolkningstilnærming.⁴²

Studerer man den effekten de strukturelle forholdene (skoler, inntektsnivå, osv.) har på utdanningsdeltagelsen, skjønner man hvordan kausalitet og seleksjons henger sammen. I et samfunn med sterkt begrenset tilgang på ressurser er det naturlig å støtte seg til kausale forklaringer. Økte ressurser til foreldrene i form av inntekt eller utdanning vil i slike tilfeller

⁴² Effekten av foreldrenes inntekt kan for eksempel skyldes seleksjon, mens effekten av antall søsken kan ha med kausalitet å gjøre.

være svært verdifull. På den annen side er det lite sannsynlig at kausale årsaker kan forklare forskjeller i utdanningsdeltagelse for individer med veldig høy tilgang til ressurser.

Koeffisientene fra modell 1 tyder på at foreldrenes inntekt og utdanning er variablene som i størst grad påvirker sannsynligheten for barnas deltagelse i høyere utdanning. I avsnitt 5.1.1 utdypet jeg hvorfor inntekt sannsynligvis ikke kan brukes til å forklare store forskjeller i dette utvalget.⁴³ Black, Devereux og Salvanes (2005) fant at høye korrelasjonene mellom foreldres og barns utdanning i utvalget de studerte, først og fremst skyldes arvede evner og ikke spillovereffekt av utdanning. Ut i fra dette er det rimelig å anta at sammenhengene mellom de viktigste forklaringsvariablene og sannsynligheten for deltagelse i høyere utdanning, for dette utvalget, i størst grad må forklares ut i fra seleksjon. Å øke foreldrenes utdanning eller inntekt vil derfor ikke direkte øke sannsynligheten for at majoriteten av utvalget tar høyere utdanning.

Det kan likevel tenkes at manglende tilgang på ressurser kan forklare hvorfor enkelte faller utenfor utdanningssystemet.⁴⁴ Barn fra familier med psykososiale problemer kan for eksempel ha tilgang på betydelig færre menneskelige og materielle ressurser enn barn fra ”vanlige” hjem. Chevalier et al., (2005) viste til at barn fra familier med små ressurser gjør det merkbart bedre når de får oppfølging fra det offentlige. Det er ikke utenkelig at dette er tilfelle også under norske forhold.

For dette utvalget, ser det ut til at seleksjon er det som i størst grad kan forklare de positive sammenhengene mellom sannsynlighet for deltagelse i høyere utdanning for barnet og viktige forklaringsvariabler som foreldrenes inntekt og utdanning. Det er likevel sannsynlig at kausale årsaker kan forklare at enkeltindivider faller utenfor. Politiske virkemidler som må settes inn for å øke det generelle utdanningsnivået i befolkningen er derfor annerledes enn de man bør benytte for å hindre at noen faller utenfor utdanningssystemet.

5.3 Grad av sosial mobilitet

I innledningen ble det introdusert at det er et politisk ønske om høy utdanningsmobilitet i befolkningen. Det er derfor naturlig å evaluere graden av sosial mobilitet når det gjelder

⁴³ Løken (2006) antyder også langt på vei at den kausale effekten av inntekt på utdanningsdeltagelsen for denne kohorten er lav. Shea (2000) fant på sin side at kausalitet i liten grad kunne brukes til å forklare hvorfor familiens inntekt har positiv sammenheng med barnas skolegang.

⁴⁴ Både Raaum (2003) og Shea (2000) viser til at for familier med minst ressurser, kan økt inntekt gi bedre utdanningsprestasjoner til barna.

utdanning for utvalget. Den sanne utdanningsmobiliteten er vanskelig å estimere siden strukturelle forhold er i kontinuerlig endring. Forskjeller mellom barn og foreldre kan derfor skyldes samfunnsmessige endringer, ikke økt mobilitet. Datamaterialet kan likevel benyttes til å si noe om overordnede sammenhenger.

Figur 3.1 viser at andelen med høyere utdanning har økt kraftig mellom generasjonene. Veksten har kommet som en følge av en storstilt kapasitetsøkning i utdanningsvesenet.⁴⁵ Etableringen av flere universiteter, høyskoler og studiefinansieringsordninger har ført til at mulighetene foreldre og barn sto overfor er svært forskjellige. For foreldregenerasjonen var nåløyet for å ta høyere utdanning svært trangt. For kohorten født i 1968 har valget i mye større grad vært basert på individuelle ønsker. De strukturelle forholdene ligger dermed til rette for en høy utdanningsmobilitet.

Tabell 3.7 og figurene 3.3-3.10 viser at det er sterke sammenhenger mellom familiens sosioøkonomiske suksess og barnas tilgang til høyere utdanning. I alle tilfeller var høy foreldreutdanning og inntekt sammenfallende med større sannsynlighet for at barna har utdanning på høyskole- eller universitetsnivå. Dette er en indikasjon på at den sosiale reproduksjonen er høy. Datamaterialet viser likevel at også en del av barna fra familier med mindre ressurser har høyere utdanning. Rekrutteringen til høyere utdanning ser derfor ut til å være bredere enn tidligere. I hvilken grad dette innbærer at utdanningsmobiliteten for denne kohorten er høy må vurderes ut i fra hvem som tar høyere utdanning som følge av kapasitetsøkningen.

Barna fra familiene med høyest sosioøkonomisk suksess er relativt lett å identifisere. Siden få av foreldrene har høyere utdanning, og inntektsfordelingene mellom familiene er jevne, er det derimot vanskeligere å kartlegge hvilke barn som kommer fra sjiktet like under de mest suksessrike familiene.⁴⁶ Hvis rekrutteringen til høyere utdanning først og fremst skjer fra denne gruppen kan man stille spørsmålsteget med hvor høy den sosiale mobiliteten egentlig er. Bredere rekruttering skyldes da først og fremst økt kapasitet.

⁴⁵ Kapasitetsøkningen beskrives blant annet i Telhaug (1994). Også statistikk fra SSB underbygger dette.

⁴⁶ Med dette mener jeg familier med relativt høy andel sosioøkonomisk suksess, hvor det er strukturelle begrensninger som gjør at foreldrene ikke har høyere utdanning.

I avsnitt 5.2 kom det frem at sammenhengene mellom variablene i de estimerte modellene, i størst grad kan forklares ut i fra seleksjon. Hvis denne tilnærmingen er korrekt, er dette en indikator på at majoriteten av førstegangsakademikerne er barn av foreldre som hadde tatt høyere utdanning under bedre strukturelle forhold. Ut i fra disse resonnementene kan man trekke slutningen om at den sosiale utdanningsmobiliteten trolig ikke er så høy som forbedringene i de strukturelle forholdene skulle tilsi.⁴⁷

⁴⁷ Vabø og Aamodt (2005) konkluderer også med at den sosiale utdanningsmobiliteten i Norge ikke er spesielt høy.

KAPITTEL 6: AVSLUTNING

6.1 Konklusjon

Effektiv politikk avhenger av kunnskap om det feltet politikken skal implementeres på. Formålet med denne oppgaven er å øke kunnskapen om hvilke mekanismer som påvirker sannsynligheten for å ta høyere utdanning. Det er derfor naturlig å avslutte oppgaven med å sette resultatene inn i en større samfunnsmessig kontekst.

De gyldige resultatene stemmer i hovedsak overens med tidligere forskning. Foreldrenes utdanning og inntekt har positiv samvariasjon med sannsynligheten for høyere utdanning, og disse ser ut til å være de viktigste kjennetegnene. Antall søsken har negativ effekt på sannsynligheten, mens mors alder har positiv effekt på sannsynligheten for deltagelse i høyere utdanning.

For å få en større innsikt i det kvalitative innholdet av resultatene, var det nødvendig å vurdere om sammenhengene skyldes kausalitet eller seleksjon. En slik evaluering legger sterke føringer for forståelsen av tematikken, og vil også prege hvilke virkemidler som tas i bruk for å nå utdanningspolitiske mål. Virkemidlene ved kausale sammenhenger er intuitive. Ved å øke mengden av forklaringsvariablene til aktuelle grupper vil man nå politiske mål. Skyldes sammenhengene seleksjon kan det være vanskeligere å utforme effektiv politikk fordi det er mer komplisert å påvirke samspillet mellom "arv og miljø".

Tradisjonelt har virkemidlene som blir benyttet for å nå utdanningspolitiske mål i Norge, bygget på en kausal tilnærming til sammenhengene. For å øke det generelle utdanningsnivået og den sosiale utdanningsmobiliteten, ble kapasiteten i høyere utdanning bygget ut og det ble etablert finansielle støtteordninger som studielån. Denne utbyggingspolitikken tok form like etter andre verdenskrig. De strukturelle forholdene for utdanning var da relativt dårlige⁴⁸, og virkemidlene har derfor gitt gode resultater. Som vi så i den deskriptive delen av oppgaven har andelen med høyere utdanning vokst kraftig, ulempen med å være kvinne i forhold til høyere utdanning har forsvunnet, og Vabø og Aamodt (2005) peker på at den negative effekten av å bo langt unna høyere utdanningsinstitusjoner er redusert.

⁴⁸ Se Telhaug (1994) for en videre diskusjon av dette.

I et grafisk bilde er det er ikke urimelig å anta at kurven for øking av det generelle utdanningsnivået på et eller annet tidspunkt vil flate ut i forhold til kurven for økte investeringer i virkemidler omkring utdanning. Spørsmålet er om dette knekkpunktet er nådd innenfor den kohorten som denne masteroppgaven er bygget på.

I avsnitt 5.2 ble det drøftet hvorvidt samvariasjonene i dette utvalget skyldes kausalitet eller seleksjon. Jeg argumentert, med bakgrunn i annen empiri, for at sammenhengene mellom familiekarakteristika og barns deltagelse i høyere utdanning i stor grad må tilskrives seleksjon. Dette kan være en indikasjon på at nettopp ovennevnte knekkpunkt er nådd. I så fall bør dette få konsekvenser for den videre utviklingen av virkemidlene for å oppnå politisk mål om stadig høyere utdanningsnivå på befolkningen.

På den andre siden ser det ut til at det er klare sammenhenger mellom utdanningsnivå og begrenset tilgang til ressurser. Derfor vil det fortsatt lønne seg å bruke tradisjonelle virkemidler for å hindre at enkeltindivider faller utenfor utdanningssystemet.

6.2 Kritikk og videre forskning

For å få frem sammenhenger og kunne generalisere, baserer økonometriske studier seg på forenklinger av virkeligheten. I noen tilfeller blir forenklingene likevel for grove, og det mislykkes å etablere sammenhenger mellom variabler slik det var tiltenkt. Å benytte mors bostedskommune i 1960 som grunnlag for avstandsmatrisen viste seg å være en for unyansert forenkling. Problemet med å benytte denne variabelen ble diskutert underveis med veileder, men det ble enighet om å fortsette og heller presentere stoffet på måten som er gjort. I ”virkelig” forskning er resultatene fra denne oppgaven lite interessante, men i en læringsprosess som en masteroppgave er ment å være, har problemet med avstandsvariabelen vært en nyttig erfaring. Praktisk i form av feilsøking, og i vurdering om resultatene kunne stemme overens med virkeligheten, men også fordi det har stimulert kritisk sans i forhold til andre empiriske studier.

Ved eventuell videre forskning hadde det selvsagt vært interessant å vite hvilke effekt avstand har på deltakelse i høyere utdanning. Siden høyere utdanningsinstitusjoner i Hordaland og Sør-Trøndelag i hovedsak befinner seg i Bergen og Trondheim kan fylkene også tjene som

et godt referansepunkt for å undersøke hvilken effekt etablering av distriktshøyskoler har hatt på deltagelsen til høyere utdanning i andre fylker.

LITTERATURLISTE

Behrman, J. & Rosenzweig, M. (2001) Does Increasing Women's Schooling Raise the Schooling of the Next Generation? **Penn Institute for Economic Research**. Working Paper 01-051.

Black, S., Devereux, P. & Salvanes, K. (2005) Why the Apple Doesn't Fall Far: Understanding the Intergenerational Transmission of Human Capital. **American Economic Review**, No 95. S. 437-449.

Caneiro, P. & Heckman, J. (2002) The Evidence on Credit Constraints in Post-Secondary Schooling. **The Economic Journal**, 112 (October) s. 705-734

Caneiro, P & Heckman, J. (2003) Human Capital Policy. I Heckman og Kreuger sin **Inequality in America: What Role for Human Capital Policy?** (Cambridge, MA: MIT Press)

Card, D. (1993) Using Geographic Variation in College Proximity to Estimate the Return to Schooling. **Industrial Relation Section Princeton University**, working paper No. 317. Tilgjengelig fra <http://www.irs.princeton.edu/pubs/pdfs/317.pdf> [Nedlastet 12. November 2006]

Chevalier, A., Harmon, C., O'Sullivan, V. & Walker, I. (2005) The Impact of Parental Income and Education on the Schooling of Their Children. **IZA Discussion Paper No. 1496**.

Dearden, L. (2004) Credit Constraints and Returns to the Marginal Learner, **Institute for Fiscal Studies**, mimeo

Frenette, M. (2002) Too Far to Go On? Distance to School and University Participation. **Statistics Canada**, Analytical Studies Research Paper Series. Tilgjengelig fra <http://www.statcan.ca/english/research/11F0019MIE/11F0019MIE2002191.pdf> [Nedlastet 12. November 2006]

Frenette, M. (2003) Access to College and University: Does Distance Matter? **Statistics Canada**, Analytical Studies Research Paper. Tilgjengelig fra <http://dsp-psd.pwgsc.gc.ca/Collection/Statcan/11F0019MIE/11F0019MIE2003201.pdf> [Nedlastet 12. November 2006]

Heckman, J. & Masterov, D. (2004) The Productivity Argument of Investing in Young Children. **Invest in Kids Working Group Committee for Economic Development**, working paper. Tilgjengelig fra http://www4.fgv.br/cps/simulador/infantil/Apresenta%C3%A7%C3%A3o/dugger_2005-07-24_dvm.pdf [Nedlastet 10. Mars 2006]

Heckman, J. (2005) Lessons From the Technology of Skill Formation. **National Bureau of Economic Research Working Paper Series**. Working paper 11142. Tilgjengelig fra <http://www.nber.org/papers/w11142>.

Hernes, G. (1976) Utdanning og Ulikhet. **NOU 1976:46**

Hægland, T. (2003) Økonomisk Avkastning av Utdanning. **Statistiske Analyser**, nr 60. Statistisk Sentralbyrå.

Lauer, C. (2003) Family Background, Cohort and Education: A French-German Comparison Based on a Multivariate Ordered Probit Model of Educational Attainment. **Labour Economics**, 10 s. 231-251

Løken, K. (2006) **The Impact of Family Income on Children's Level of Education: Causation or Selection?** Akademisk avhandling, Universitetet i Bergen.

Raaum, O. (2003) Familiebakgrunn, Oppvekstmiljø og Utdanningskarrierer. **Statistiske Analyser**, nr 60. Statistisk Sentralbyrå

Romer, D. (2006) **Advanced Macroeconomics**. McGraw-Hill Irwin. 3rd e.

Shea, J. (2000) Does parents' money matter? **Journal of Public Economics**, 70 (2) s. 155-184.

Telhaug, A. (1994) **Norsk Skoleutvikling Etter 1945**. Oslo, Didakta Norsk Forlag.

Todd, P & Wolpin, K. (2003) On the Specification and Estimation of the Production Function for Cognitive Achievement. **The Economic Journal**, 113 (February) s. F3-F33.

Vabø, A. & Aamodt, P. (2005) Kvalitetsreformen og Universitetene som Masseutdanningsinstitusjoner **Norsk Institutt for Studier av Forskning og Utdanning / Senter for Innovasjonsforskning**, Skriftserie 2/2005. Tilgjengelig fra <http://nifu.pdc.no/publ/pdf/2005/S-2005-2.pdf> [Nedlastet 10. januar 2007]

Wooldridge, J. (2003) **Introductory Econometrics: A Modern Approach, 2e**. Thomson South-Western.

Aamodt, P. (1982) Utdanning og Sosial Bakgrunn. Statistisk Sentralbyrå. **Samfunnsøkonomiske Studier (SØS)**, nr. 51.

Internettkilder:

Hjemmesiden til Statistisk Sentralbyrå <http://www.ssb.no>

Hjemmesiden til Gule Sider <http://www.telefonkatalogen.no>

Hjemmesiden til Kunnskapsdepartementet <http://odin.dep.no/kd/>

APPENDIKS

Appendiks A, Økonometri

Resonnement for at en binær modell angir sannsynlighet:

Hvis modellen er riktig spesifisert, vil forventningsverdien til feilledet, u , gitt alle verdier av de uavhengige variablene være null.

$$E(u | x_1, \dots, x_k) = 0$$

Forventningsverdien av y gitt forklaringsvariablene er derfor

$$E(y | x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Siden y er en binær variabel som enten er null eller en, vil sannsynligheten for at dummyen er lik 1, være den samme som forventningsverdien til y .

$$P(y=1 | x_1, \dots, x_k) = E(y | x_1, \dots, x_k)$$

Innsatt for $E(y | x_1, \dots, x_k)$

$$P(y=1 | x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

På grunn av dette angir den binære modellen sannsynligheten for at den avhengige variabelen = 1.

Resonnement for at lineære sannsynlighetsmodeller alltid har heteroskedastisitet:

Når y er en binær variabel vil variansen avhengig av \mathbf{x} være

$$\text{Var}(y | \mathbf{x}) = p(\mathbf{x})[1-p(\mathbf{x})]$$

$p(\mathbf{x})$ er forkortelse for $p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$

Dette innebærer at så lenge sannsynligheten avhenger av minst en av de uavhengige variablene så må det eksistere heteroskedastisitet.

Appendiks B, Regresjonsresultater

Regresjonsresultater

Resultater fra den ekte Probitestimeringen av modell 1, som viser generelle sammenhenger.

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Permanentinntekt	0,4654	0,0847	0,4389	0,0888
Mors utdanningsnivå	0,2305	0,0249	0,2747	0,0254
Fars utdanningsnivå	0,2347	0,0191	0,2239	0,0190
Antall barn i familien	-0,1078	0,0238	-0,1144	0,0236
Mors alder	0,0279	0,0045	0,0234	0,0045

Alle koeffisientene er skarpt bestemt.

Resultater fra Modell 1, hele kohorten.

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Permanentinntekt	0,1044	0,0080	0,0826	0,0084
Mors utdanningsnivå	0,0792	0,0035	0,0968	0,0035
Fars utdanningsnivå	0,0834	0,0027	0,0723	0,0028
Antall barn i familien	-0,0289	0,0028	-0,0366	0,0031
Mors alder	0,0082	0,0006	0,0097	0,0006

Alle koeffisientene er skarpt bestemt.

Resultater fra når reisetid benyttes som mål på avstand

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Reisetid	0,0004	0,0002	0,0003	0,0002
Permanentinntekt	0,1425	0,2182	0,1383	0,0239
Mors utdanningsnivå	0,0848	0,0087	0,097	0,0085
Fars utdanningsnivå	0,0731	0,0064	0,0809	0,0065
Antall barn i familien	-0,0337	0,0071	-0,0373	0,0075
Mors alder	0,0092	0,0014	0,0081	0,0015

Reisetid er statistisk signifikant på 5 %-nivå for menn og 10 %-nivå for kvinner.

Resten av koeffisientene er skarpt bestemt.

Resultater fra når også individer fra Sogn og Fjordane og Nord-Trøndelag inkluderes.

Høyere utdanning	Menn		Kvinner	
Variabel	Koeffisient	St.avvik	Koeffisient	St.avvik
Avstand	0,0002	0,0001	0,0003	0,0001
Permanentinntekt	0,1339	0,0230	0,0956	0,2237
Mors utdanningsnivå	0,0691	0,0092	0,1042	0,0090
Fars utdanningsnivå	0,0762	0,0069	0,7267	0,0069
Antall barn i familien	-0,0448	0,0071	-0,0420	0,0079
Mors alder	0,0110	0,0014	0,0080	0,0015

Alle resultatene er skarpt bestemte bortsett fra mennenes avstandsvariabel som er statistisk signifikant på 1%-nivå.

Appendiks C, Grafer

Inntektsfordeling til utvalget. Gjennomsnittelige deflaterte permanentinntekter.

