
Intergenerasjonell inntektsmobilitet: hvordan forklare den geografiske variasjonen?

Sverre Falck Langaas

Masteroppgave

Masteroppgaven er levert for å fullføre graden

Master i samfunnsøkonomi

Universitetet i Bergen, Institutt for økonomi

Desember 2020



UNIVERSITETET I BERGEN

Forord

Jeg vil først og fremst takke min veileder Espen Bratberg som har gitt mange konstruktive tilbakemeldinger underveis i prosessen med å skrive denne oppgaven. Jeg vil også takke Trond Pedersen ved Norsk senter for forskningsdata (NSD) for raske og hjelpsomme tilbakemeldinger når jeg har hatt spørsmål om plattformen microdata.no. Til slutt vil jeg takke de tidligere masterstudentene Peder Grimstad Helset og Magnus Pettersen Nytnun for at de la ved utdrag fra sine microdata-script. Dette var hjelpsomt da jeg skulle begynne å lage mine egne script. Jeg har valgt å legge ut mine egne script, både for at det skal være mulig å reprodusere analysen og forhåpentligvis for at fremtidige studenter som skal bruke microdata i deres oppgaver kan ha nytte av det.¹ Utformingen av scriptene er også beskrevet i vedlegget.

Oppgaven er skrevet i L^AT_EX, figurene er laget med Matplotlib i Python og kartene er laget med programvaren QGIS. Databehandlingen og den statistiske analysen er utført i microdata.no. Gjennom denne plattformen har jeg hatt tilgang til en stor mengde registerdata uten å måtte søke om å få data utlevert. Nedsiden er at det ikke er mulig å se på individdataene og at det er færre funksjonaliteter enn i vanlige statistiske programvarer. Det har vært noe tungvint og teknisk krevende å gjennomføre denne analysen i microdata, men hovedinntrykket av plattformen har vært veldig positivt og det har ikke lagt vesentlige begrensinger på analysene jeg har ønsket å gjennomføre.

¹<https://github.com/SverreFL/masteroppgave>

Sammendrag

Jeg bruker administrative data til å dokumentere geografiske forskjeller i intergenerasjonell inntektsmobilitet mellom ulike steder i Norge. Deretter gjennomfører jeg et kvasieksperiment for å vurdere den relative betydningen til ulike forklaringer på de dokumenterte forskjellene i inntektsmobilitet. Jeg undersøker om barn som flytter til kommuner med høyere absolutt mobilitet – i betydningen av at barna som vokser opp hos foreldre med tilsvarende inntekt i gjennomsnitt havner høyere i den nasjonale inntektsfordelingen – også selv oppnår høyere inntekt. Siden valg om hvor familien flytter er endogent vil enkle sammenligninger medføre seleksjonskjevhet. Den identifiserende antagelsen er at seleksjonsskjevheten er konstant for barn som flytter i ulike aldre. Den kausale eksponerings-effekten kan da isoleres ved å sammenligne utfall til barn som flytter i ulike aldre. Dette forskningsdesignet ble først benyttet av Chetty og Hendren (2018) på amerikanske data. I likhet med deres studie finner jeg at barna som flytter til steder med høyere absolutt mobilitet også selv oppnår høyere inntekt og at effekten er større for barn som flytter tidligere i oppveksten. Resultatet viser at den geografiske variasjonen i absolutt mobilitet delvis kan forklares med forskjeller i oppvekstmiljø på ulike steder, og dette kan settes i sammenheng med litteratur om nabolageffekter.

Innhold

1 Innledning	1
2 Empirisk litteratur	4
2.1 Elastisitet og korrelasjon	4
2.2 Overgangsmatriser og rangmobilitet	6
3 Teori og kausale mekanismer	9
3.1 Becker-Tomes modellen	9
3.2 Forklaringer på forskjeller i inntektsmobilitet innad i land	10
3.3 Eksperiment for å vurdere relativ betydning av forklaringer	10
4 Empirisk rammeverk	13
4.1 Rangmobilitet	13
4.2 Forskningsdesign og identifiserende antagelse	15
4.2.1 Definisjon av eksponeringseffekt	16
4.2.2 Estimere eksponeringseffekt med observasjonsdata	17
5 Data og utvalg	20
5.1 Definisjon av kjerneutvalget	20
5.2 Definisjon av flytteutvalgene	21
5.3 Definisjon av inntekt	23
6 Nasjonale estimat	25
6.1 Intergenerasjonell inntektselastisitet	25
6.2 Inntektsrangering	26
7 Geografisk variasjon i inntektsmobilitet	29
7.1 Geografiske enheter	29
7.2 Relativ mobilitet	30
7.3 Absolutt mobilitet	30

8	Eksponeeringseffekt av oppvekstmiljø	37
8.1	Forskningsdesign	37
8.2	Estimat av eksponeeringseffekt	38
8.3	Validering av forskningsdesignet	43
8.4	Sammendrag	44
9	Konklusjoner	46
A	Appendiks	51
A.1	Klassisk målefeil	58
A.2	Utforming av script i microdata.no	60
A.2.1	Deskriptiv	60
A.2.2	Flyttedelen	61

Figurer

6.1	Forhold mellom inntekt til foreldre og deres barn	26
6.2	Forhold mellom inntektsrangering til far og barn	27
7.1	Relativ mobilitet	31
7.2	Absolutt mobilitet i Stavanger/Sandnes og Trondheim	32
7.3	Absolutt mobilitet	36
8.1	Estimat på b_m (1)	39
8.2	Estimat på b_m (2)	40
A.1	Estimat på b_m med inntekt ved ulike aldre	54
A.2	Absolutt mobilitet i øverste halvdel av inntektsfordelingen	55
A.3	IGE-estimat med ulike mål på barns inntekt	56
A.4	Korrelasjon av inntektsrangering med ulike mål på inntekt	57

Tabeller

5.1	Utvalgsbeskrivelse	21
5.2	Antall flyttere	22
5.3	Inntekt i kjerneutvalget	24
6.1	Nasjonal overgangsmatrise	28
8.1	Estimat på gjennomsnittlig eksponeringseffekt	41
A.1	Økonomiske regioner	51
A.2	Askeladd-indeks	53

Kapittel 1

Innledning

Mulighet til å bevege seg oppover den sosiale rangstigen regnes ofte som en god egenskap ved et samfunn. Det betyr at det er mulig å komme seg opp og frem gjennom personlige egenskaper som evner og innsats. Alternativet er at muligheter blir bestemt av familiebakgrunn. Da er det flaks i fødselslotteriet som bestemmer barns muligheter og slik at sosial posisjon går i arv gjennom generasjoner. En grunn til at sosial mobilitet regnes som et gode er hensynet til rettferdighet. Ulikhet i utfall som skyldes egne valg og personlige egenskaper oppfattes som mer rettferdig enn det som skyldes flaks. Lav sosial mobilitet kan også være ineffektivt i den grad det skyldes at posisjoner går i arv gjennom nettverk, i stedet for å tilfalle de mest egnede kandidatene.

I motsetning til likhet i utfall er det derfor tilsynelatende en politisk konsensus om at likhet i muligheter er et gode. Denne politiske konsensusen er kun overfladisk og opprettholdes av at begrepet *like muligheter* har flere ulike betydninger (Swift 2014). I svak form er det samsvarende med meritokrati og betyr at posisjoner bør gå til de meste kompetente kandidatene, noe som utelukker nepotisme og diskriminering. Dette vil det kunne være stor grad av enighet om, men for mange vil det være også være for svakt til å medføre reell likhet av muligheter. I sterk form betyr det også at personer har like muligheter til å tilegne seg kompetansen. Dette er et langt mer krevende konsept. En fullstendig utjevning av muligheter vil legge betydelig restriksjoner på friheten familier har til å hjelpe egne barn som vil være lite spiselig for mange. I praksis er det derfor politiske uenigheter om balansen mellom like muligheter og andre goder. Selv om det ikke er konsensus om fullstendig utjevning av muligheter, i hvert fall ikke i sterk form, så er det stor grad av enighet om å sikre at alle barn har muligheter uavhengig av familiebakgrunn. Dette er en av begrunnelsene for velferdsstaten.

Sosial mobilitet betyr bevegelse på den sosiale rangstigen. Begrepet er knyttet til idéen om klassereise og bevegelse mellom ulike sosiale lag. Det eksisterer en omfattende litteratur som forsøker å tallfeste grad av sosial mobilitet ved å se på sammenheng mellom målbare utfall til foreldre og deres barn. I praksis studeres ofte sammenhengen mellom inntekten til foreldre og deres barn, der høy korrelasjon medfører høy persistens av utfall og dermed

lav sosial mobilitet.¹ Studier viser at de nordiske landene skiller seg ut ved høy grad av sosial mobilitet, mens det blant vestlige land er lavest sosial mobilitet i Storbritannia og USA (f.eks. Jäntti mfl. 2006).

Slike sammenligninger mellom ulike land er informative, men de ulike målene på sosial mobilitet kan likevel undervurdere den faktiske rigiditeten i samfunnet. Inntekt kan være en unøyaktig mål for plassering på den sosial rangstigen fordi jobber har ulike egenskaper som gjør de attraktive og lønn kun utgjør én dimensjon. Det kan være jobber som har høy lønn for å kompensere for at de er lite attraktive av andre grunner. Tilsvarende kan det være slik at barn som vokser opp i familier med høy inntekt opplever større frihet til å velge usikre og til dels dårlige betalte jobber som er attraktive av andre grunner, slik som å være kunstner eller andre kreative yrker. Det kan derfor tenkes at persistens av inntekt mellom generasjoner er lavere enn persistens av underliggende muligheter.

De nordiske landene skiller seg ikke bare ut ved å være mer mobile, men også ved å ha bedre tilgang på administrative data med inntektsmål for hele populasjonen. Det er først ganske nylig at slike data har blitt tilgjengelige i USA. Raj Chetty har sammen med ulike medforfattere har sammen med ulike medforfattere gjennomført en rekke banebrytende analyser av sosial mobilitet med utgangspunkt i dette nye datamaterialet. I Chetty mfl. (2014) dokumenter de betydelig geografisk variasjon i sosial mobilitet innad i landet, der stedene med høyest sosial mobilitet har estimat som ligner på mobiliteten i de nordiske landene. Den geografiske variasjonen medfører at barn som vokser opp med foreldre med samme inntekt i gjennomsnitt havner ulike steder på den nasjonale inntektsfordelingen avhengig av hvor de vokser opp. Chetty m. fl. (2018) gjør tilsvarende analyse helt ned på nabolagsnivå og viser at det også er betydelig lokal variasjon innad i *counties*.²

I tillegg til å dokumentere den geografiske variasjonen i sosial mobilitet forsøker de også å forklare hva som forårsaker variasjonen. I Chetty m. fl (2014) gjennomfører de en korrelasjonsanalyse og finner noe overraskende at sosial mobilitet i liten grad er korrelert med egenskaper ved det lokale arbeidsmarkedet som sysselsetting og lønnsvekst. Chetty og Hendren (2018) gjennomfører et kvasiexperiment som utnytter variasjon i alderen barna når de flytter til å undersøke den relative betydning av forskjeller i egenskaper ved arbeidsmarked, oppvekstmiljø og befolkning på forskjeller i sosial mobilitet. De finner igjen at forskjeller i egenskaper ved arbeidsmarkedet har liten forklaringskraft siden det er liten effekt av å flytte til et sted med høyere sosial mobilitet i alderen like før barna går inn i arbeidsmarkedet. Effekten avhenger lineært med antall år med eksponering før barna er 23 år, og de konkluderer med at mye av den geografiske variasjon i sosial mobilitet i stor grad kan forklares med forskjeller i oppvekstmiljø.

Det er mulig å knytte analysen til litteraturen om betydningen av nabolag på barns

¹Sammenhengen mellom inntekt til foreldre og barn betegnes heretter som inntektsmobilitet. Høy persistens av inntekt mellom generasjoner medfører lav inntektsmobilitet.

²Tallmateriale er tilgjengelig som et søkbart atlas på <https://www.opportunityatlas.org/>

utfall. Observasjonelle studier har dokumentert betydelig variasjon i barns langssiktige økonomisk utfall avhengig av i hvilke nabolag de vokser opp (f.eks. Sampson mfl. 2002, Sharky og Faber 2014). På en annen side har eksperimentelle studier i liten grad klart å dokumentere en kausal effekt av nabolag på barns utfall (f.eks. Oreopoulos 2003, Ludwig mfl. 2013). I kvasiexperimentet benyttes en grovere geografisk inndeling enn nabolag, men i den grad det kan dokumenteres en kausal eksponeringseffekt for de ulike stedene kan dette tolkes som bevis for at en kausal effekt ved oppvekstmiljø i nabolagene på de ulike stedene. For å anvende dette i utforming av policy gjenstår det å finne ut gjennom hvilke mekanismer nabolag påvirker barnas utfall, og deretter forsøker å forbedre nabolagene med lav sosial mobilitet. Det første steget er uansett å dokumentere en kausal effekt.

Jeg skal gjennomføre en lignende analyse på norske data. De geografiske forskjellene i inntektsmobilitet er tidligere dokumentert med samme metode og datagrunnlag av Norberg-Schulz og Østhus (2018). Jeg bruker en annen utvalgsavgrensing og mål på inntekt, men hovedkonklusjonene er de samme. I likhet med deres studie finner jeg at det er mindre geografisk variasjon i sosial mobilitet i Norge enn i USA og at barna som er oppvokst på Vestlandet har størst sannsynlighet for å bevege seg oppover den sosiale rangstigen. Analysen av kvasiexperimentet på norske data er et originalt bidrag. Jeg finner i likhet med Chetty og Hendren (2018) at barn som flytter til steder med høyere mobilitet også selv oppnår høyere inntekt og at effekten er større for barn som flytter tidlig i oppveksten.

Strukturen av oppgaven er som følger. I andre kapittel presenterer jeg ulike mål på inntektsmobilitet og gir en kort oversikt over den empiriske litteraturen. I tredje kapittel beskriver jeg en modell som brukes til å forklare forskjeller i inntektsmobilitet mellom ulike land og endringer over tid, og jeg argumenterer for at lignende mekanismer kan forklare den geografiske variasjon i inntektsmobilitet innad i land. I fjerde kapittel forklarer jeg forskningsdesignet. Deretter beskriver jeg utvalg og variabeldefinisjoner i kapittel 5, presenterer nasjonale estimat i kapittel 6, dokumenter geografisk variasjon i kapittel 7 og analyserer kvasiexperimentet i kapittel 8. Til slutt forsøker jeg å komme med noen konklusjoner.

Kapittel 2

Empirisk litteratur

Det eksisterer en omfattende empirisk litteratur med studier som undersøker sammenhengen mellom inntekten til foreldre og deres barn. De tidlige studiene brukte i stor grad intergenerasjonell inntektselastisitet (IGE) eller intergenerasjonell inntektskorrelasjon som begge reduserer sammenhengen til et enkelt tall. Dette gjør det mulig å sammenligne grad av inntektsmobilitet mellom ulike land og se på endringer over tid. På en annen side kan ikke fange opp alle aspekter ved sammenhengen slik at det nødvendigvis er et tap av informasjon. Det er derfor blitt utviklet andre metoder som gjør det mulig å studere ikke-linearitet og heterogenitet i sammenhengen innad i land. Det eksisterer ulike mål som måler ulike ting, og de har ulike styrker og ulemper. Jeg skal forsøke å gi en kort oversikt over denne litteraturen.

2.1 Elastisitet og korrelasjon

Den intergenerasjonelle inntektselastisiteten er et sammendragsmål angir den forventede prosentvise økningen i barns inntekt hvis foreldrenes inntekt øker med én prosent. Det er det mulig å estimere fra ligningen

$$\log(y_i) = \alpha + \beta \log(x_i) + \epsilon_i, \quad (2.1)$$

der y_i er inntekt til barnet, x_i er inntekt til foreldre og β er IGE. For at parameteren skal kunne tolkes som elastisiteten må vi anta at sammenhengen er lineær slik at elastisiteten ikke avhenger av inntekten til foreldre, $E[\epsilon_i | \log(x_i)] = 0$. Høyere elastisitet medfører større persistens av inntekt mellom generasjonene og dermed lavere inntektsmobilitet. Elastisiteten kan øke både fordi korrelasjonen mellom inntekt til foreldre og deres barn

øker, men også som følge av endringer i inntektsulikheten. Dette følger av at

$$\begin{aligned}\beta &= \frac{\sigma_{xy}}{\sigma_x^2} \\ \implies \beta \frac{\sigma_x}{\sigma_y} &= \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \rho,\end{aligned}\tag{2.2}$$

der σ_y er standardavviket til logaritmen av inntektene til barna, σ_x er tilsvarende for foreldre og ρ korrelasjonen mellom inntekten til foreldre og deres barn. Den intergenerasjonelle inntektskorrelasjonen ρ er et alternativt mål på sammenhengen. Hvis $\rho = 1$ er barnas inntekt fullstendig bestemt av inntekt til foreldrene slik at det er ingen inntektsmobilitet.

I praksis er det betydelige utfordringer ved å oppnå presise estimat av disse parameterne. Årsaken er at inntektsmålene vi er interessert i er de *permanent inntektene* til foreldre og deres barn. Den permanente inntekten kan for eksempel betraktes som den gjennomsnittlige inntekten i yrkesaktiv alder. Hvis vi i stedet bruker inntekt i enkeltår som proxy for den permanente inntekten vil det være avvik mellom størrelsene vi observerer og størrelsene vi egentlig er interessert i. For det første vil det være en livssyklus-kjevhet som følge av at inntekter har tendens til å følge en konkav løpebane over yrkesløpet. Hvis inntekt måles for tidlig eller for sent vil det systematisk avvike fra permanent inntekt. Det er også en utfordring at inntekten har ulik løpebane for ulike yrkesgrupper og stagnerer tidligere i mange lavinntektsyrker. Haider og Solon (2006) viser at hvis inntekt til barna måles for tidlig vil dette medføre at estimatene på IGE overvurderer grad av inntektsmobilitet. For det andre vil det også være svingninger i inntekt over yrkesløpet. Dette kan modelleres som såkalt *klassisk målefeil* og medfører at estimatene på IGE blir for lave slik at grad av inntektsmobilitet blir overvurdert. Solon (1992) og Zimmerman (1992) demonstrer at målefeilene medfører at IGE i USA sannsynligvis er rundt 0.4, noe som er betydelig høyere enn estimatene på rundt 0.2 i oversiktsartikkelen til Becker og Tomes (1986). Senere argumenterer Mazumder (2005) for at IGE i USA er rundt 0.6. Dette illustrerer spredningen av estimatene i litteraturen.

Det er usikkerhet knyttet til de eksakte målene på IGE i ulike land. Målefeil gjør estimatene inkonsistente og estimatene er sensitive for definisjoner av inntekt, behandling av observasjon uten inntekt og avgrensninger av utvalget. På tross av disse utfordringene er det mulig å avdekke klare empiriske mønster ved å sammenligne estimat fra studier i ulike land som bruker lignende metodologi. For det første er det en tendens til at land med høyere inntektsulikhet har lavere inntektsmobilitet. Denne sammenhengen har blitt kjent som *The Great Gatsby Curve* (Krueger 2012, Corak 2013). Videre er det slik at land med større forskjeller i gjennomsnittsinntekt til personer med og uten universitetsgrad har lavere inntektsmobilitet (Corak 2013). Jäntti mfl. (2006) er eksempel på en studie som gjør det mulig å sammenligne estimat i ulike land. De estimerer IGE i de nordiske

landene, Storbritannia og USA med lignende utvalg og variabeldefinisjoner. Hvis estimatene justeres for målefeil elastisiteten er rundt 0.5-0.6 i USA, rundt 0.3 i Storbritannia og under 0.3 i alle nordiske landene (Black og Devereux 2010).

Jäntti mfl. (2006) finner også at forksjellene i IGE i stor grad skyldes ulik mobilitet i endepunktene av inntektsfordelingen til foreldre. I USA er det større persistens i endepunktene, slik at barn som vokser opp i lavinntektsfamilier har større sannsynlighet for også å ha lavinntekt og tilsvarende for barn av foreldre med veldig høy inntekt. Barn som vokser opp med foreldre mellom disse ytterpunktene har lignende inntektsmobilitet i de ulike landene. Dette illustrerer en svakhet ved IGE og korrelasjon som mål på inntektsmobilitet. De reduserer sammenhengen mellom inntekten til foreldre og deres barn til ett enkelt tall. Dette gjør de velegnet til å sammenligne mobilitet mellom land, men de kan ikke brukes til å avdekke ikke-lineariteter i sammenhengen. Videre er det lite egnet til å studere heterogenitet i inntektsmobilitet for ulike grupper innad i land.

Både intergenerasjonell inntektselastisitet og korrelasjon er mål på det Chetty mfl. (2014) betegner som relativ mobilitet. Det sier noe om hvor høy inntekt barn som vokser opp i familier med lav inntekt har i forhold til barn som vokser opp i familier med høyere inntekt. Den relative mobiliteten kan dermed øke både som følge av at barn i lavinntektsfamilier får bedre utfall og at barn fra høyinntektsfamilier får dårligere utfall. Det er dermed ikke entydig at økning i relativ mobilitet er et gode. Absolutt mobilitet, derimot, gir mål på gjennomsnittlig utfall til barn som vokser opp med foreldre på et bestemt sted i inntektsfordelingen. Jeg skal nå presentere alternativ mål på inntektsmobilitet som i tillegg til å beskrive relativ mobilitet også kan karakterisere den absolutte mobiliteten.

2.2 Overgangsmatriser og rangmobilitet

De alternative målene tar utgangspunkt i foreldrene og barnas plassering på inntektsfordelingen i sin kohort. I stedet for å se direkte på inntekt blir foreldre og deres barn plassert i kvantiler av den kumulative inntektsfordelingen i deres kohort. Det er da mulig å undersøke sammenhengen mellom kvantilen til foreldre og deres barn. Hvis det er fullstendig mobilitet slik at foreldrenes inntekt ikke har noen betydning for barnas inntekt vil det være like stor sannsynlighet for bevegelse mellom alle kvantilene. Hvis det er ingen mobilitet vil alle barna ha inntekt i samme kvantil som sine foreldre. Jäntti m. fl. (2006) deler inntektsfordelingen inn i kvintiler og lager en overgangssmatrise som for hver kvantil i inntektsfordelinge til foreldrene viser sannsynligheten for at barn som vokser opp med foreldre i den kvintilen havner i de ulike kvintilene av inntektsfordelingen i egen kohort. Dette gjør det mulig å studere ikke-linearitet i inntektsmobilitet. De finner at det i USA er større persistens i endepunktene av inntektsfordelingen enn i de andre landene. Det er altså større sannsynlighet for at barna som vokser opp med foreldre i den laveste kvintilen også får inntekt i lavest kvantil som voksen. Tilsvarende er det slik at de med foreldre i

den høyest kvintilen i USA har større sannsynlighet for å forbli der enn i de andre landene i studien. For barn som vokser opp i de andre kvintilene er transisjonsmatrisene relativt like i de ulike landene og de konkluderer derfor at forskjellene i estimatene på intergenerasjonell inntektselastisitet i stor grad skyldes forskjeller i persistens i endepunktene av inntektsfordelingen.

Overgangsmatriser gir en mer detaljert beskrivelse av simultanfordelingen mellom utfallene til barn og deres foreldre. Samtidig gjør den detaljerte beskrivelsen det vanskeligere å sammenligne mellom ulike land og se på endringer over tid. Hvis inntektsfordelingen er delt inn i kvintiler blir det for eksempel 25 celler i matrisen. Det er derfor utviklet ulike sammendragsmål for informasjonen i transisjonsmatrisen. Det er for eksempel vanlig se på andelen av dem som vokser opp med foreldre i laveste kvintil som havner i den høyeste kvintilen. Chetty m. fl. (2014) betegner dette som andelen som oppnår *den amerikanske drømmen*.¹

Bruken av overgangsmatriser har blitt kritisert av Bhattacharya og Mazumder (2008) fordi det innebærer en vilkårlig partisjonering av inntektsfordelingen og kun ser på bevegelse mellom de vilkårlige inndelingene. De foreslår en alternativ tilnærming der inntektsfordelingen blir delt inn i prosentiler og bruker andelen av barn som havner i høyere prosentil enn sine foreldre blir et mål på inntektsmobilitet. Dette legger større vekt på små bevegelser som ikke blir fanget opp av den grovere inndelingen i for eksempel kvintiler. Dahl og DeLeire (2008) beskriver sammenhengen mellom prosentil til foreldre og deres barn med ligningen

$$y_i = \alpha + \psi p_i + \epsilon_i, \quad (2.3)$$

der y_i er prosentil til barn og p_i er prosentil til foreldrene og $cov(\epsilon_i, p_i) = 0$. De viser at sammenhengen er omtrent lineær i amerikanske data slik at

$$E[y|p] \approx \alpha + \psi p. \quad (2.4)$$

Denne ligningen kan brukes som mål på absolutt mobilitet for barn som vokser opp med foreldre i prosentil p . Samtidig er helningskoeffisienten også korrelasjonen mellom prosentil til barn og deres foreldre,

$$\psi = \frac{cov(y, p)}{var(p)} = \frac{cov(y, p)}{sd(p)sd(y)} = corr(y, p), \quad (2.5)$$

slik at det kan brukes som mål på relativ mobilitet. Det er denne metodologien Chetty og hans medforfattere bruker til å dokumentere og analysere geografisk variasjon innad i USA. Jeg skal bruke samme metodologi til å gjøre lignende analyse på norske data og

¹Jeg velger i likhet med Aasland (2019) å betegne det som Askeladd-indeksen.

metodologien er derfor nærmere beskrevet nedenfor.

Bratberg m. fl. (2015) bruker lignende metodologi på data fra Norge, Sverige, Tyskland og USA. De finner mindre geografisk variasjon på grovere regional inndeling. Norberg-Schulz og Østhus (2018) bruke samme metodologi som studien til Chetty m. fl (2014) og finner betydelige forskjeller i mobilitet på tvers av kommuner. De finner at forskjellene i gjennomsnittlige utfall til barn som vokser opp i lavinntektsfamilier i stor grad gjenspeiler regionale forskjeller i inntektsvekst. Det er både høy inntektsvekst og mobilitet på vestlandet og de ser dette i sammenheng med oljeindustrien. Vektleggingen av betydningen til arbeidsmarkedet er i kontrast med konklusjon til Chetty m. fl (2014) som fant at inntektsmobilitet i liten grad er korrelert med egenskaper til arbeidsmarkedet.

Kapittel 3

Teori og kausale mekanismer

Den empiriske litteraturen har dokumentert betydelig geografisk variasjon i inntektsmobilitet. Det er både variasjon mellom ulike land og innad i land, og både mellom ulike regioner og helt ned på nabolagsnivå (Chetty mfl. 2018). Samtidig er det en voksende litteratur som forsøker å forklare de observerte forskjellene og forstå gjennom hvilke mekanismer posisjonen på den sosiale rangstigen blir overført gjennom generasjoner (Black & Devereux 2010). I denne delen skal jeg gi en kort skisse av litteraturen og bruke det teoretiske rammeverket til å diskutere hvilke mekanismer som kan forklare geografisk variasjon i inntektsmobilitet innad i land.

3.1 Becker-Tomes modellen

Denne litteraturen har utgangspunkt i Becker-Tomes modellen (1979, 1986) der foreldre påvirker barns utfall gjennom overføringer av humankapital. Det kan både være passive investeringer gjennom overføring av genetiske egenskaper og aktive investeringer. Barna får deretter avkastning på sin beholdning av humankapital i form av høyere lønn på arbeidsmarkedet. Posisjonen på den sosiale rangstigen kan dermed gå i arv gjennom generasjoner ved at foreldre med høy inntekt investerer mer i sine barns humankapital. Det eksisterer en omfattende empirisk litteratur om faktorer som påvirker barns langsiktige utfall, og i dette rammeverket kan disse faktorene - enten det er barnets kosthold eller høytlesning på sengen - betraktes som investeringer i barnets humankapital.

Solon (2004) har utviklet modellen for å gjøre den mer egnet til å forklare variasjon i intergenerasjonell inntektsalitet mellom ulike land og endringer over tid. I denne modellen kan det forventede utfallet til et barn forstås som et samspill mellom familiebakgrunn, arbeidsmarked og politiske virkemidler. I modellen vil avkastning på humankapital i arbeidsmarkedet sentralt for å forklare variasjon i mobilitet, både fordi påvirker incentivene til privat investering og fordi den resulterende økonomiske ulikheten medfører ulik kapasitet til å investere. Dette forsterker sammenhengen mellom inntekten til foreldre og inntekten barna oppnår som voksne, og reduserer dermed inntektsmobiliteten. Det kan

dempes gjennom progressive offentlige investeringer i barns humankapital som reduserer betydningen av private investeringer og dermed familiebakgrunn.

Denne modellen kan i stor grad forklare forskjellene i intergenerasjonell inntektselastisitet (IGE) mellom ulike land, siden det er høyere elastisitet i land med høyere økonomisk ulikhet og større forskjell i gjennomsnittslønn til personer med og uten universitetsutdannelse (Corak 2013).¹ Det er noen unntak fra dette mønsteret, blant annet Italia som både har høy IGE og lav premium på universitetsutdannelse. Dette kan tyde på at foreldre der i større grad kan overføre sin sosiale posisjon gjennom andre mekanismer enn investering i humankapital.

3.2 Forklaringer på forskjeller i inntektsmobilitet innad i land

Dette rammeverket kan også benyttes til å undersøke hvilke mekanismer som forårsaker den observerte geografiske variasjonen i inntektsmobilitet innad i land. I et rent meritokrati er det rimelig at investering i humankapital er mekanismen som forklarer sammenheng mellom inntekt til foreldre og deres barn. Jeg tenker at det da vil være tre mulige forklaringer på hvorfor barn som vokser opp i familier med tilsvarende inntekt i gjennomsnitt oppnår ulik inntekt som voksne avhengig av hvor de vokser opp. For det første kan det være slik at familier som er bosatt på ulike steder er systematisk forskjellige og investerer ulike mengder i barnas humankapital betinget av samme inntekt. En annen mulig forklaring er at det skyldes egenskaper ved de lokale arbeidsmarkedene som gjør at barn med samme beholdning av humankapital i gjennomsnitt oppnår ulik lønn. Det vil også være sammenheng mellom disse to forklaringene siden avkastningen på humankapital påvirker incentivene til å investere, i tråd med Becker-Tomes-Solon-modellen over. Den tredje forklaringen er at stedet barna vokser opp påvirker beholdningen av humankapital. Det medfører at barn som i utgangspunktet vokser opp i identiske familier oppnår ulik beholdning av humankapital og dermed ulik lønn avhengig av hvor de vokser opp.

3.3 Eksperiment for å vurdere relativ betydning av forklaringer

De ulike forklaringene er langt fra gjensidig utelukkende og virkeligheten er alltid mer sammensatt. Vi kan likevel tenke på hvordan et hypotetiske eksperiment kunne blitt

¹Corak betegner denne forskjellen som *premium* på universitetsutdannelse og det kan tolkes som et mål på avkastningen av humankapital i arbeidsmarkedet. Det påvirker foreldrenes incentiv til å investere i barnas humankapital, og dermed familiens betydning barnas beholdning av humankapital og lønn de oppnår som voksne. I modellen er dette den sentrale mekanismen som forklarer forskjeller i sosial mobilitet, og sammenhengen mellom inntektsulikhet og sosial mobilitet kan tolkes i lys av dette.

gjennomført for å vurdere den relative betydningen til de ulike forklaringene. Anta at vi kan trekke et tilfeldig utvalg av familier som blir flyttet fra et sted med lav absolutt mobilitet til et sted med høy absolutt mobilitet for barn som vokser opp med foreldre i en gitt inntekt. Hvis forskjellene i absolutt mobilitet utelukkende skyldes at familier som er bosatt ulike steder er forskjellige på måter som påvirker barnas humankapital betinget av inntekt vil ikke flyttingen påvirke barnas gjennomsnittlige utfall, utover eventuelle kostnader ved selve flyttingen.² Hvis forskjellene i mobilitet skyldes forskjeller i egenskaper ved arbeidsmarkedene vil vi forvente en positiv behandlingseffekt på barna som blir flyttet. Alderen til barna når de flytter vil kun påvirke effekten av behandlingen i den grad det påvirker sannsynligheten for å bli værende på det nye stedet og delta i det lokale arbeidsmarkedet. Hvis forskjellene derimot skyldes kausal effekt ved oppvekstmiljøet som påvirker beholdning av humankapital vil vi forvente at størrelsen på behandlingseffekten avhenger av antall år med eksponering for det nye oppvekstmiljøet.

Lignende eksperiment har blitt gjennomført. Det største og mest ambisiøse er Moving to Opportunity (MTO) som ble gjennomført mellom 1994-1998 i USA. Samlet var det 4600 familier fra lavinntektsområder i som deltok, og der et tilfeldig utvalg ble tildelt en kupong som subsidierte leiekostnader i et rikere nabolag. Hensikten var å undersøke om barna som flytter vekk fra nabolag med konsentrert fattigdom oppnår bedre utfall som voksne. Tidlige analyser av dette eksperimentet klarte ikke å påvise noen positiv kausal effekt på barnas inntekt av å bli flyttet til et rikere nabolag, og finner tvert imot en svak negativ effekt (Kling mfl. 2007). De konkluderte derfor med at nabolag i liten grad påvirker barns inntekt som voksne. Chetty, Hendren og Katz (2016) argumenter for at dette skyldes at de tidlige studiene så på utfallene til barn som allerede var ungdommer da de flyttet siden de ikke hadde data om inntekt til de yngre barna. De gjør en ny analyse av eksperimentet og finner en positiv kausal effekt på inntekten til barn som flyttet før de var 13 år. Dette er konsistent med at det er kumulativ eksponering for bedre oppvekstmiljø som bestemmer den kausale effekten av flytting. Den negative effekten for de eldre barna kan delvis forklares ved at flyttekostnadene utveier eksponeringseffekten.

I den grad det er kausale stedeffekter som påvirker barnas utfall gjenstår det å undersøke gjennom hvilke mekanismer det skjer for å anvende denne innsikten til å forbedre utfallene på steder med lav mobilitet, slik at barn har mulighet til å bevege seg oppover den sosiale rangstigen uavhengig av familiebakgrunn og hvor de vokser opp. Hvis det skyldes eksternaliteter av humankapital gjennom såkalte *sosiale interkasjonseffekter*³ vil det avhenge av befolknings sammensetningen på ulike steder og det kan være vanskelig å

²For eksempel ved at det tar tid å tilpasse seg det nye miljøet og bygge opp sosialt nettverk. For enkelthets skyld vil heretter se bort ifra slike kostnader i det hypotetiske eksperiment. For å eliminere denne effekten kan man eventuelt anta at flyttekostnader ikke avhenger av hvor barna flytter og sammenligne utfall mellom barn som flytter til steder med ulik absolutt mobilitet, i stedet for å sammenligne med utfall til barna som blir igjen.

³Det som på engelsk kalles peer effects.

påvirke dette gjennom politiske virkemidler. Hvis det derimot skyldes egenskaper ved for eksempel skoler, fritidstilbud eller lignende vil det være større handlingsrom. Dette ligger utenfor rammen av oppgaven. Jeg skal nå utvikle et empirisk rammeverk som gjør det mulig å både dokumentere geografisk variasjon i inntektsmobilitet, samt å gjennomføre et kvasiexperiment for å undersøke om det også er en slike kausale stedeffekter i Norge.

Kapittel 4

Empirisk rammeverk

Formålet med oppgaven er å dokumentere geografisk variasjon i inntektsmobilitet og undersøke om det eksisterer en kausal effekt på barns inntekt av å bli eksponert for *bedre* oppvekstmiljø. Et bedre oppvekstmiljø betyr i denne sammenhengen at barna som vokser opp med foreldre med tilsvarende inntekt i gjennomsnitt har høyere inntekt som voksne. Jeg vil i likhet med Chetty mfl. (2014) og Chetty og Hendren (2018) benytte rangmobilitet til å beskrive sammenhengen mellom inntekt til foreldre og deres barn på ulike steder. I denne seksjonen vil jeg først beskrive denne metodologien. Deretter vil jeg gi en oversikt over forskningsdesignet Chetty og Hendren (2018) bruker til å identifisere den kausale eksponeringseffekten fra observasjonsdata fra barn som flytter. Jeg vil både presentere resultatet fra deres studie og de ulike strategiene de benytter for å vurdere troverdigheten til den identifiserende antagelsen. Til slutt vil jeg mer formelt definere parameteren jeg vil estimere.

4.1 Rangmobilitet

Rangmobilitet gir et fleksibelt rammeverk som gjør det mulig å sammenligne mobilitet i delpopulasjoner og som imotsetning til inntektselastisitet og korrelasjon ikke trenger å anta en lineær sammenheng. Videre er det et mer robust mål som ikke er like sensitiv for behandling av observasjoner uten inntekt og målefeil som følge av begrenset antall år med observert inntekt (jmf. figur A.3 og A.4 i appendikset, samt Chetty mfl. 2014).

I stedet for å se direkte på inntekt blir foreldre og deres barn delt inn i prosentiler av inntektsfordelingen til sin respektive kohort. Dette gjør det mulig å se på sammenhengen mellom plassering på inntektsfordelingen til foreldre og deres barn. La prosentil til barn i være y_i og prosentil til foreldrene være p_i . Et mål på sammenhengen er gitt ved den betingede forventningsfunksjonen $E[y|p]$ som for hver prosentil angir den gjennomsnittlige plasseringen på inntektsfordelingen til barn som vokser opp med foreldre i den prosentilen. Dette kan estimeres ved å ta de korresponderende gjennomsnittene i utvalget. Sammenhengen kan da visualiseres med et såkalt *binned scatterplot* som gjør det mulig å

undersøke ikke-lineariter i sammenhengen. Dette er et ikke-parametrisk estimat på $E[y|p]$ som ikke krever noen antagelser om funksjonell form.

Videre er det også mulig å estimere sammenhengen for ulike delpopulasjoner, for eksempel inndelt etter geografi, innvandrerstatus eller rase. Det gjør det mulig å betrakte forskjeller inntektsmobilitet for ulike grupper. La c_i være en variabel som indikerer hvilken gruppe observasjon i tilhører. Da kan vi på tilsvarende måte estimere $E[y|p, c]$. Dette kan vi bruke til å sammenligne absolutt mobilitet i de ulike gruppene. Hvis

$$E[y_i|p_i = p, c_i = j] > E[y_i|p_i = p, c_i = k] \quad (4.1)$$

vil gruppe j ha større absolutt mobilitet enn gruppe k for barn som vokser opp med foreldre i den gitte prosentilen fordi de i gjennomsnitt havner lenger opp i den nasjonale inntektsfordelingen. Problemet med bruke dette til å rangere steder etter grad av absolutt mobilitet er at det avhenger av hvilken prosentil i foreldrenes inntektsfordeling vi bruker. Det vil da også bare bruke informasjon fra barna med foreldre i den prosentilen gitte slik det blir sensitivt for enkeltobservasjoner.

Chetty mfl. (2014) viser at sammenhengen mellom foreldrenes og prosentil og den gjennomsnittlige prosentil til barna deres er omtrent lineær og at denne egenskapen er robust for de ulike geografiske områdene. Dette er et rent empirisk resultat og stemmer ikke like bra på norske data (jmf. figur 6.2) ettersom kurven er brattere på toppen og bunnen av inntektsfordelingen.¹ Jeg velger likevel å følge Chetty mfl. (2014) og Norberg-Schulz og Østhus (2018) som parametriserer den betingede forventningsfunksjonen som

$$\bar{y}_{pc} = E[y|p, c] = \alpha_c + \psi_c p. \quad (4.2)$$

Disse parametrene kan estimeres separat for hvert geografisk område c med minste kvadratsmetode

$$y_i = \alpha_c + \psi_c p_i + \epsilon_i, \quad (4.3)$$

der y_i og p_i er prosentiler i den nasjonale inntektsfordelingen. For å rangere de geografiske områdene fra lavest til høyest absolutt mobilitet er det fortsatt nødvendig å redusere sammenhengen til et enkelt tallmål. Chetty mfl. (2014) argumenterer for at det gjennomsnittlige utfallet til barn som vokser opp med foreldre i den 25. prosentilen er et godt mål siden det med en lineær sammenheng tilsvarer det gjennomsnittlige utfallet til barn som vokser opp med foreldre i nedre halvdel av inntektsfordelingen. Dette målet på absolutt

¹Sammenhengen ser ut til å være godt beskrevet av en kubisk funksjon. Det kunne nok ha vært en god idé å gjøre en sensitivitetsanalyse med den parametrisering av sammenhengen.

mobilitet er gitt ved

$$\bar{y}_{25,c} = \alpha_c + \psi_c \cdot 25. \quad (4.4)$$

4.2 Forskningsdesign og identifiserende antagelse

Forskningsdesignet går ut på å bruke data fra barn som flytter i ulike aldre av oppveksten til å tilnærme det ideelle eksperimentet der et tilfeldig utvalg av familier blir flyttet. Dette gjør det mulig å undersøke om barn som flytter til steder med høyere absolutt mobilitet for barn av foreldre med tilsvarende inntekt også selv oppnår bedre utfall som voksne. Utfordringen med å bruke observasjonsdata til å estimere kausal effekt er at valget om å flytte er endogen. Familiene som flytter kan da være systematisk forskjellige fra de som blir igjen. Det kan også være slik at familier som flytter til *bedre* steder er systematisk forskjellige fra familiene som flytter *dårligere steder*.² Enkle sammenligninger av utfallene til barna som flytter og de som blir igjen, eller mellom de som flytter til steder med ulik absolutt mobilitet, vil da fange opp denne seleksjonsskjevheten.

Den identifiserende antagelsen som gjør det mulig å isolere den kausale effekten av å bli eksponert for et *bedre* sted er at alderen barnet flytter i er så godt som tilfeldig bestemt. Dette betyr at gitt at familiene flytter til samme sted, så vil det ikke være noen systematiske forskjeller mellom familier som flytter når barna er i ulike aldre. Det vil da være mulig å kvantifisere den kausale effekten av eksponering ved å sammenligne utfall til barn som flytter i ulik alder. Dette er en svakere antagelse enn at valg om å flytte og valg av destinasjon er tilfeldig bestemt, men det er fortsatt en forholdsvis sterk antagelse. Chetty og Hendren har flere ulike strategier for å vurdere troverdigheten til antagelsen.

Det kan for eksempel være slik at familier som flytter når barne er yngre er systematisk forskjellig på måter som påvirker barnas inntekt, for eksempel ved å ha høyere utdannelse. Dette er det mulig å kontrollere for ved å sammenligne utfall til søsken som er i ulik alder når familien flytter. Siden søsknene vokser opp med samme foreldre vil dette kontrollere for alle egenskapene til foreldre som er konstante over tid. En annen utfordring er at det kan være andre egenskaper ved oppvekstmiljøet som endrer seg samtidig med flyttingen. Beslutningen om å flytte kan ofte ha sammenheng med endring i samlivsstatus til foreldre, et formuesjokk eller at foreldre får ny jobb. Det kan da være slik at forskjellene i utfallene til søsken skyldes ulik eksponering for andre egenskaper ved oppvekstmiljøet som endret seg samtidig med at familien flyttet. De forsøker å kontrollere for dette ved å la effekt av endring i foreldrenes samlivsstatus og inntekt variere med alderen barna flytter.

Chetty og Hendren finner at estimatene er robuste for de ulike spesifikasjonene. De

²*Bedre* og *dårligere* steder brukes i betydningen av at barna som vokser opp med foreldre i en gitt prosentil av den nasjonale inntektsforeldre i gjennomsnitt havner henholdsvis høyere eller lavere på den nasjonale inntektsfordelingen enn barna som vokser opp hos foreldre med tilsvarende inntekt på stedet det sammenlignes med.

finner at utfallene til barna som flytter konvergerer mot de gjennomsnittlige utfallene til barna som vokser opp der og at graden av konvergering avhenger omtrent lineært med antall år med eksponering. For hvert år barna tilbringer på et sted der barna med foreldre med samme inntekt havner én prosentil høyere i inntektsfordelingen vil barnas forventede prosentil øke med 0.04 prosentiler. De finner at denne effekten varer frem til barna er rundt 23 år og at det deretter ikke er noen effekt av å flytte. De konkluderer derfor med at forskjellene i inntektsmobilitet mellom ulike steder i stor grad skyldes kausal effekt av oppvekstmiljøene på ulike steder og at det er kumulativ eksponering som bestemmer størrelse på behandlingseffekten.

For å underbygge at disse resultatene faktisk skyldes kausale stedseffekter utnytter de heterogenitet i inntektsmobilitet mellom ulike grupper på ulike steder. Det er for eksempel slik at steder med mye kriminalitet har dårligere utfall for gutter enn jenter. De finner at for familier som flytter til disse områdene så konvergerer utfallene til barna mot utfallene til sitt kjønn, slik at sønnene får dårligere utfall. Graden av konvergens avhenger også her proposjonalt med kumulativ eksponering. Videre ser de også på andre utfall enn inntekt, slik som sannsynlighet for å gjennomføre universitetsgrad eller å bli gravid som ungdom. De finner akkurat det samme mønsteret med konvergens til utfallene til populasjonen i destinasjonen som avhenger proposjonalt med antall år med eksponering.

Disse resultatene underbygger at det er kausale stedseffekter som påvirker utfallene til barna som blir eksponert for de ulike oppvekstmiljøene. De konkluderer med at virkemidler som forbedrer oppvekstmiljøet, for eksempel gjennom investering i skoler, er bedre egnet til å øke inntektsmobilitet enn virkemidler som tar sikte på å styrke det lokale arbeidsmarkedet. Jeg skal undersøke om det også er mulig å på slike kausale stedseffekter i Norge. På grunn av mindre utvalg og mindre heterogenitet i inntektsmobilitet vil det ikke være mulig å gjennomføre strategiene Chetty og Hendren benytter for å vurdere troverdigheten til den identifiserende antagelsen. Det er da betryggende at de får omtrent samme estimat med alle de ulike spesifikasjonen de bruker. Jeg skal nå formelt definere parametrene som jeg skal estimere.

4.2.1 Definisjon av eksponeringseffekt

Formålet er å undersøke om det eksisterer kausale stedseffekter som påvirker utfall til barn. Forskningsdesignet utnytter at barn flytter til *bedre nabolag* i ulike aldre slik at de får ulik eksponering. Rangmobilitet kan brukes til å beskrive hvordan de gjennomsnittlige utfallene til barn avhenger av foreldrenes inntekt på ulike steder. Det gjennomsnittlige prosentilen i den nasjonale inntektsfordelingen til barn som vokser opp på sted d hos foreldre med inntekt p kan skrives som

$$\bar{y}_{pd} = \alpha_d + \psi_d p. \tag{4.5}$$

Anta nå at det er mulig å gjennomføre et eksperiment der et tilfeldig utvalg av barn på et gitt sted blir flyttet til ulike steder. Det vil da være mulig å se på sammenhengen mellom barnas utfall og grad av inntektsmobilitet på stedene de blir flyttet til, og hvordan denne sammenhengen avhenger av alderen til barnet når det blir flyttet. Er det slik at barn som blir flyttet til nabolag med høyere mobilitet får bedre utfall? Vil det kun være effekt for barn som flyttet før en viss alder? Dette er spørsmål som dette hypotetiske eksperimentet vil kunne gi svar på. Mer konkret vil det være mulig å kvantifisere den gjennomsnittlige kausale effekten av å bli flyttet i alder m til et sted der barn som vokser opp med foreldre i samme prosentil har én prosentil bedre utfall i gjennomsnitt. Vi kan estimere ligningen

$$y_i = \alpha + \beta_m \bar{y}_{pd} + \theta_i, \quad (4.6)$$

der θ_i fanger opp effekten av familiens investeringer i humankapital og andre variabler som påvirker barnas utfall. I et tilfeldig eksperiment vil $cov(y_{pd}, \theta_i) = 0$ slik slik at β_m kan estimeres konsistent med minste kvadrats metode. Parameteren β_m representerer den gjennomsnittlige kausale effekten på barnas prosentil i den nasjonale inntektsfordelingen av å tilbringe årene fra alder m og utover på sted med én prosentil bedre gjennomsnittlige utfall. Ved å estimere denne effekten for barns som flytter i ulike aldre er det mulig å definere eksponeringseffekten ved alder m som

$$\gamma_m = \beta_m - \beta_{m+1}. \quad (4.7)$$

Slik Chetty og Hendren argumenterer for er det det interessant å estimere eksponeringseffektene $\{\gamma_m\}$ av flere grunner. For det første vil $\gamma_m > 0$ for noen m indikere at sted har kausal betydning for utfall til barn for noen aldre, noe som ikke er et trivielt resultat i lys av at flere tidligere eksperimentelle studier ikke har påvist kausal effekt. For det andre vil $\{\gamma_m\}$ gi informasjon om hvilke år av oppveksten der stedeffekter har størst betydning for utfall til barn. Videre vil den kumulative eksponeringseffekten over hele oppvekste, $\beta_0 \equiv \sum_{t=0}^T \gamma_m$, være et mål på hvilken andel av differanse i utfall som skyldes kausal effekt av sted. Hvis $\beta_0 = 0$ impliserer det at hele forskjellen skyldes seleksjon og at flytting av barn ikke har noen effekt på deres utfall. Hvis $\beta_0 = 1$ impliserer det at det ikke er noen seleksjonseffekt og at forskjellene i utfall utelukkende skyldes ulik eksponering for ulike stedeffekter.

4.2.2 Estimere eksponeringseffekt med observasjonsdata

Det er mulig å tilnærme seg dette hypotetiske eksperimentet ved å bruke ved å bruke data om inntekt til barn i familier som av ulike grunner velger å flytte én gang under oppveksten til barna. Det er i hovedsak to utfordringer med å estimere parametrene i (4.7) på observasjonsdata. For det første vil valg av destinasjon være endogen slik at det

kan korrelere med andre variabler som påvirker barns utfall. Da er ikke lenger betingelsen om at $cov(y_{pd}, \theta_i) = 0$ oppfylt slik at minste kvadrats metode vil være en inkonsistent estimator for β_m . Denne estimatoren vil konvergere i sannsynlighet til

$$b_m = \frac{cov(y_{pd}, y_i)}{var(y_{pd})} = \frac{cov(y_{pd}, \alpha_m + \beta_m y_{pd} + \theta_i)}{var(y_{pd})} \quad (4.8)$$

$$= \beta_m + \delta_m, \quad (4.9)$$

der $\delta_m = \frac{cov(y_{pd}, \theta_i)}{var(y_{pd})}$ er seleksjonseffekten som fanger opp effekten av andre faktorer som påvirker utfall til barna og som er korrelert med utfallene i destinasjonen som familien velger å flytte til. Den identifiserende antagelsen som gjør det mulig å estimere den kausale eksponeringseffekten er at seleksjonseffekten ikke av alderen til barna når familien flytter, $\delta_m = \delta$ for alle m . Med andre ord er det slik at *hvor* barna flytter kan korrelere med andre variabler som påvirker deres utfall, men betinget av hvor de flytter så antar vi at *alderen* de flytter ikke er korrelert med de utelatte variablene. Da vil

$$b_m - b_{m+1} = \beta_m + \delta - (\beta_{m+1} + \delta) = \gamma_m, \quad (4.10)$$

slik at differansen i effekt av å flytte ved ulike aldre isolerer den kausale eksponeringseffekten.

Den andre utfordringen er at for å få et tilstrekkelig stort utvalg må vi bruke barn som flytter *fra* ulike steder. Hensikten er fortsatt å undersøke om det har kausal effekt å flytte til sted som har bedre utfall enn opprinselsen. Det er derfor nødvendig å kontrollere for de gjennomsnittlige utfallene på stedene barna flytter fra. På samme måte som over kan de gjennomsnittlige utfallene til barn som vokser opp på sted o med foreldre i prosentil p av den nasjonale inntektsfordelingen skrives som

$$\bar{y}_{po} = \alpha_o + \psi_o p. \quad (4.11)$$

En observasjonell analog til (4.6) kan da være

$$y_i = \alpha^1 + \alpha^2 \bar{y}_{po} + b_m \Delta_{pod} + \epsilon_{1i}, \quad (4.12)$$

der $\Delta_{pod} := \bar{y}_{pd} - \bar{y}_{po}$. Ved å kontrollere for gjennomsnittlig utfall vil b_m fortsatt kunne tolkes som effekt av å flytte ved alder m til en destinasjon der de permanentene innbyggerne med foreldre i samme prosentil i gjennomsnitt havner én prosentil høyere i den nasjonale inntektsfordelingen. Spesifikasjonen kan utvides til å estimere effekt ved å flytte i ulike aldre ved

$$y_i = \alpha^1 + \alpha^2 \bar{y}_{po} + \sum_m b_m I(m_i = m) \Delta_{pod} + \kappa \Delta_{pod} + \epsilon_{2i}. \quad (4.13)$$

I likhet med Chetty og Hendren velger jeg også å kontrollere for at såkalte forstyrrelseskostnader ved å flytte kan avhenge av alder barnet flytter samt inntekt til foreldre. Jeg modellerer dette med en fast effekt for flyttealder samt et interkasjonsledd mellom flytteralder og prosentil p_i av inntekten til foreldre. Dette medfører at

$$\begin{aligned}
y_i = & \alpha^1 + \alpha^2 \bar{y}_{po} + \sum_m b_m I(m_i = m) (\zeta_m^1 + \zeta_m^2 p_i) \\
& + \sum_m b_m I(m_i = m) \Delta_{pod} + \kappa \Delta_{pod} + \epsilon_{3i},
\end{aligned} \tag{4.14}$$

som også er spesifikasjonen jeg vil benytte til å estimere $\{b_m\}$.

Kapittel 5

Data og utvalg

Datagrunnlaget for analysen er basert på flere administrative registre fra Statistisk sentralbyrå som er gjort tilgjengelig gjennom plattformen *microdata.no*. På denne plattformen er det mulig å koble sammen barn med deres foreldre, og det er blant annet informasjon om bosted og ulike mål på inntekt. Jeg vil først beskrive konstruksjonen av kjerneutvalget som brukes til å dokumentere geografisk variasjon i inntektsmobilitet. Deretter vil jeg beskrive utvalgene som brukes til å estimere den kausale eksponeringseffekten. Til slutt vil definere målene på inntekt som blir brukt i analysen.

5.1 Definisjon av kjerneutvalget

Utformingen av kjerneutvalget er gjort med hensyn på å få gode mål på gjennomsnittlig livsløpsinntekt til barn og foreldre. Kjerneutvalget består av individer som (1) er født mellom 1972-1980, (2) har kjent bostedskommune når de er 16 år, (3) er bosatt i Norge i alderen 31-35 år og (4) kan kobles til en far som er bosatt i Norge og har registrert positiv inntekt mellom 1993-1998. Det er også avgrenset ved at faren er født mellom 1940-1958 som sikrer at inntekten til faren er målt mellom alderen 35-58 år.¹

I kjerneutvalget registrerer jeg kun bosted når barna er 16 år. Dette blir regnet som oppvekstkommunen til barna og kan avvike fra bosted til barna når jeg registrerer inntekten deres som voksne. Jeg har valgt å bruke fars inntekt som mål på inntekten til foreldrene.² Avgrensningene i utvalget gjør at gjennomsnittsalder til far ved barnets fødsel og ved målt inntekt varierer mellom de ulike kullene (jmf. tabell 5.1). Dette er en

¹Inntektsvariabelen er tilgjengelig for årene 1993-2015. Norberg-Schulz og Østhus (2018) gjør tilsvarende analyse på samme datagrunnlag, men benytter et utvalg bestående av barn født i perioden 1978-1985. De måler inntekt til barna når de er 28-32 år (og 28-30 år for det yngste kullet siden de også kun har inntektsdata fram til 2015).

²Dette valget er i hovedsak begrunnet av begrensinger ved programvaren. Når foreldre blir koblet på datasettet med barna blir inntekten duplisert for hver av barna. Hvis jeg deretter finner kvantiler av denne variabelen vil det ikke samsvare med prosentilene av inntektsfordelingen til foreldre. Jeg tar derfor kvantiler før foreldre blir koblet på barna og dette medfører at jeg enten må se på inntekt til far eller mor.

konsekvens av begrensningene ved datagrunnlaget og gjør det vanskelig å se på endringer i inntektsmobilitet mellom de ulike kullene i utvalget, men jeg tror ikke det har stor betydning for estimatene for hele utvalget.

Tabell 5.1: Utvalgsbeskrivelse

Kohort	Antall barn ^a	... som kan kobles til far ^b	Andel ^c	Fars alder ved fødsel	Fars alder i 1993
1972	61535	43530	70.74%	26.11	47.11
1973	59030	44166	74.82%	26.53	46.53
1974	57651	45630	79.15%	26.93	45.93
1975	54400	44857	82.46%	27.22	45.22
1976	51842	43916	84.71%	27.59	44.59
1977	49649	42938	86.48%	28.03	44.03
1978	50504	44074	87.27%	28.51	43.51
1979	50442	44207	87.64%	28.88	42.88
1980	50295	43480	86.45%	29.39	42.39
Totalt	485341	396809	81.76 %	27.68	44.70

^a Antall barn som er bosatt i Norge når de er 16 år når jeg registrerer bostedskommune samt i alderen 30-35 når inntekt blir målt.

^b Antall som kan kobles til far som er født mellom 1940-1958 og bosatt i Norge i årene 1993-1998 når inntekt blir målt.

^c Andelen av barna som kan kobles på far. Forskjellen i andeler skyldes i hovedsak at større andel av foreldrene i de eldste kullene er født før 1940. For kullet i 1972 ekskluderes for eksempel fedre som var eldre enn 32 år ved barnets fødsel. Dette medfører også at gjennomsnittsalder ved fødsel øker for de yngre kullene, ettersom de eldre fedrene i mindre grad blir ekskludert fra utvalget.

5.2 Definisjon av flyttestvalgene

I flyttestvalgene benytter jeg tre ulike avgrensinger av fødselsår til barna. Årsaken til dette er at ved å måle inntekt ved tidligere alder kan jeg inkludere yngre kull og observere barn som flytter tidligere i oppveksten. På en annen side gir det dårligere mål på livsløpsinntekt til barna. Chetty og Hendren (2018) viser at estimatene ikke er sensitiv for alderen de måler inntekt. De forklarer dette med at estimatene uansett gir et mål på i hvilken grad inntekten til innflytterne konvergerer mot inntekten til de permanente innbyggerne ved den gitte alderen, og de velger derfor å benytte inntekt i alderen 24 år for å kunne observere barn som flytter så tidlig som mulig. Det er ikke tydelig at det samme argumentet er gyldig for mine data.³ Jeg velger derfor å presentere estimat med inntekt målt ved 27, 29 og 31 år. Utvalgene består av individer som (1) er født i Norge mellom 1977 og det siste året jeg kan observere inntekt ved angitt alder, (2) er bosatt i en av de 150 største kommunene

³ Dette kan skyldes større usikkerhet til estimatene slik at de er forskjellig med inntekt målt i ulik alder på grunn av støy. Det kan også skyldes at barna kommer senere ut i arbeidslivet i Norge slik at inntekt ved ung alder gir dårligere mål på barnas inntekt i Norge enn i USA (Aakvik mfl. 2012, se også figur A.4 som viser at korrelasjon av inntektsrangering til far og barn er sensitivt for alder inntekt til barn blir målt).

frem til de er 18 år, (3) er bosatt i Norge ved alderen jeg målet inntekt og (4) som kan kobles til en far som har registrert positiv inntekt mellom 1993-1998.⁴

Tabell 5.2: Antall flyttere

Flyttealder	Inntekt ved alder		
	27 år	29 år	31 år
4 år	641	-	-
5 år	1042	-	-
6 år	1455	500	-
7 år	1616	803	-
8 år	1492	898	344
9 år	1741	1134	600
10 år	1883	1353	830
11 år	1924	1439	985
12 år	2036	1567	1155
13 år	2289	1809	1426
14 år	2245	1773	1463
15 år	2334	1853	1529
16 år	3350	2867	2309
17 år	3574	3100	2519
Totalt	27638	19114	13155

Noter: Utvalget består av barn som er født fra 1977 til det seneste kullet jeg kan observere inntekt ved den angitte alderen i 2015. Flytterne består av barna som har endret folkeregistrert adresse én gang i de årene av oppveksten jeg observerer fra 1992 og fram til de er 18 år. Hvis adressen er ulik i begynnelsen år t og år $t + 1$ regner jeg at de flyttet i år t og flyttealderen er da alder til barnet i år t .

Utvalget består av to deler: de permanente innbyggere og flyttere. De permanente innbyggerne består av individene som har samme bostedskommune fra den første observasjonen i 1992 og helt fram til individet er 18 år. For den eldste kohorten som er født i 1976 betyr dette at de regnes av permanente innbyggere av den bostedskommunen de bodde i fra de var 16-18 år. Mangelen på data om bosted tidligere i oppveksten er et svakhet ved datagrunnlaget. De permanente innbyggerne brukes til å estimere $\bar{y}_{pc} = \alpha_c + \psi_c p$ for hver av de 150 kommunene. Dette brukes som mål på de gjennomsnittlige utfallene til barn med far i prosentil p av den nasjonale inntektsfordelingen som tilbringer hele oppveksten i kommune c .

Flytterne består av individene som endrer bostedskommune én gang mellom første observasjonen i 1992 og fram til individet er 18 år.⁵ Jeg regner at barna flytter i år t hvis

⁴Størrelse på kommunene er rangert etter befolkning i 2002. Jeg observerer kun bosted fra 1992 og kan kun sikre at individer er bosatt i de angitte kommunene i de årene jeg observerer. Jeg benytter kvantiler fra den nasjonale inntektsfordelingen som er beregnet før jeg avgrensar til de 150 største kommunene. I likhet med i kjerneutvalget må faren være bosatt i Norge i hele intervallet inntekten måles. Faren må være født mellom 1940-1963 som sikrer at inntekten er målt mellom alderen 30-58 år.

⁵Chetty og Hendren (2018) observerer flyttinger helt frem til barna i 30 år. I utvalget mitt vil denne

bostedskommunen i folkeregisteret er forskjellig i begynnelsen av år $t + 1$ og t . For et barn født i år t_0 er alder m ved flyttetidspunktet da $t - t_0$. Tabell 5.2 viser antallet barn som flytter med inntekt målt ved de ulike aldre. Størrelsen på utvalget er sensitiv for alderen jeg målet inntekt fordi ved å måle inntekt tidligere kan jeg inkludere yngre kull, og jeg observerer de yngre kullene i større del av oppveksten slik at det er flere tidspunkt de kan flytte på.

5.3 Definisjon av inntekt

I teorien ønsker vi å se på sammenhengen mellom den *permanente inntekten* til foreldre og deres barn. Den permanente inntekten kan for eksempel konseptualiseres som gjennomsnittlig inntekt i yrkesaktiv alder. I praksis må vi finne en observerbar proxy for denne størrelsen. Dette betyr at det er nødvendig å gjøre valg om hvilke inntekter som skal inkluderes og når de skal måles. Målet er at størrelsen vi observerer blir et godt mål på gjennomsnittsinntekten i yrkesaktiv alder slik at estimatene beskriver den sammenhengen vi er interessert i.

Jeg følger Norberg-Schulz og Østhus (2018) og definerer inntekt i hovedanalysen som summen av lønnsinntekt, netto næringsinntekt, kapitalinntekt og skattepliktige overføringer.⁶ Dette målet på inntekt bruker jeg både til å dokumentere den geografisk variasjon i inntektsmobilitet og i analysen av kvasiexperimentet. I appendikset benytter jeg pensjonsgivende inntekt som tilgjengelig tilbake til 1967 for å undersøke hvordan målene på inntektsmobilitet avhenger av hvilke aldre inntekten til far og barn blir målt. Det viser at både IGE og korrelasjonen mellom inntektsrangering er sensitiv for alderen inntekten til barna blir målt (jmf. figur A.3 og A.4). Begge målene vokser ganske raskt med alderen inntekten til barna blir målt frem til de er 33 år og flater deretter ut. Det viser også at målet på IGE er langt mer sensitiv for både alderen inntekt til far blir målt i samt lengden på intervallet.

Avgrensingen av utvalget er gjort med hensyn til at inntektsmålene jeg vil bruke kun er tilgjengelig for perioden 1993-2015. Inntektsmålet til barna er gjennomsnittsinntekten når de er 31-35 år. For foreldrene måles inntekt som gjennomsnitt av fars inntekt i perioden 1993-1998. Barna vil være i alderen 13-21 år i det første året inntekten til far måles. Dette betyr at for de eldste kohortene måles foreldrenes inntekt sent i deres oppvekst og gjerne etter de har flyttet ut.⁷ Jeg mener dette ikke er et problem siden vi uansett er interessert i sammenhengen mellom livsløpsinntekt til far og barn.

avgrensingen medføre en betydelig reduksjon av antallet barn som flytter tidlig i oppveksten siden en stor andel av de som flytter én gang før de er 18 også flytter minst én gang til før de er 30 år. Jeg velger derfor å fokusere på barna som flytter tidligere i oppveksten.

⁶Alle monetære størrelser benyttet i analysen måles i faste 2015-priser, justert for inflasjon ved hjelp av konsumprisindeksen (KPI).

⁷Til sammenligning bruker Norberg-Schulz og Østhus (2018) og Chetty mfl (2014) foreldrenes inntekt når barna er henholdsvis 12-16 år og 15-20 år.

I store deler av analysen benytter jeg kvantiler av den nasjonale inntektsfordelingen i stedet for å se på inntekt direkte. Kvantilene til fars inntekt er beregnet før observasjonene blir koblet på datasettet med barna, så det inkluderer inntekten til menn som oppfyller de andre kriteriene for å være inkludert i utvalget bortsett fra at de ikke kan kobles til et barn som oppfyller kriteriene.⁸

Tabell 5.3: Inntekt i kjerneutvalget

Inntekt	Barna	Far
Gjennomsnitt	456 000	445 000
Standardavvik	196 000	220 000
Min	74 000	108 000
25. prosentil	332 000	316 000
50. prosentil	430 000	395 000
75. prosentil	543 000	508 000
Maks	1 131 000	1 379 000

Noter: Inntekt til barna og far er målt som gjennomsnitt av inntekten i henholdsvis alderen 31-35 år og i årene 1993-1998. Inntektsmålet er definert som summen av lønnsinntekt, netto næringsinntekt, kapitalinntekt og skattepliktige overføringer. Min og maks tilsvarener nederste og øverste prosentil. Alle tallene er avrundet til nærmeste tusen kroner.

⁸Alternativet ville vært å ta kvantiler etter å ha koblet fedrene på datasettet med barna. I dette datasettet er barna koblingsnøkkel og inntekten til far blir da duplisert for hvert av barna, slik at kvantilene av denne variabelen avviker fra kvantilene av inntektsfordelingene til fedrene.

Kapittel 6

Nasjonale estimat

Jeg vil først presentere nasjonale mål på intergenerasjonell inntektsmobilitet. Deretter vil jeg presentere mål på inntektsmobilitet som tar utgangspunkt i sammenhengen mellom rangering til barn og foreldre på den nasjonale inntektsfordelingen. Disse målene er interessante i seg selv og brukes til å sammenligne graden av mobilitet i dette utvalget med andre studier på norske data, samt studier fra andre land. Det gir også et sammenligningsgrunnlag for målene på mobilitet for ulike geografiske enheter innad i landet.

6.1 Intergenerasjonell inntektselastisitet

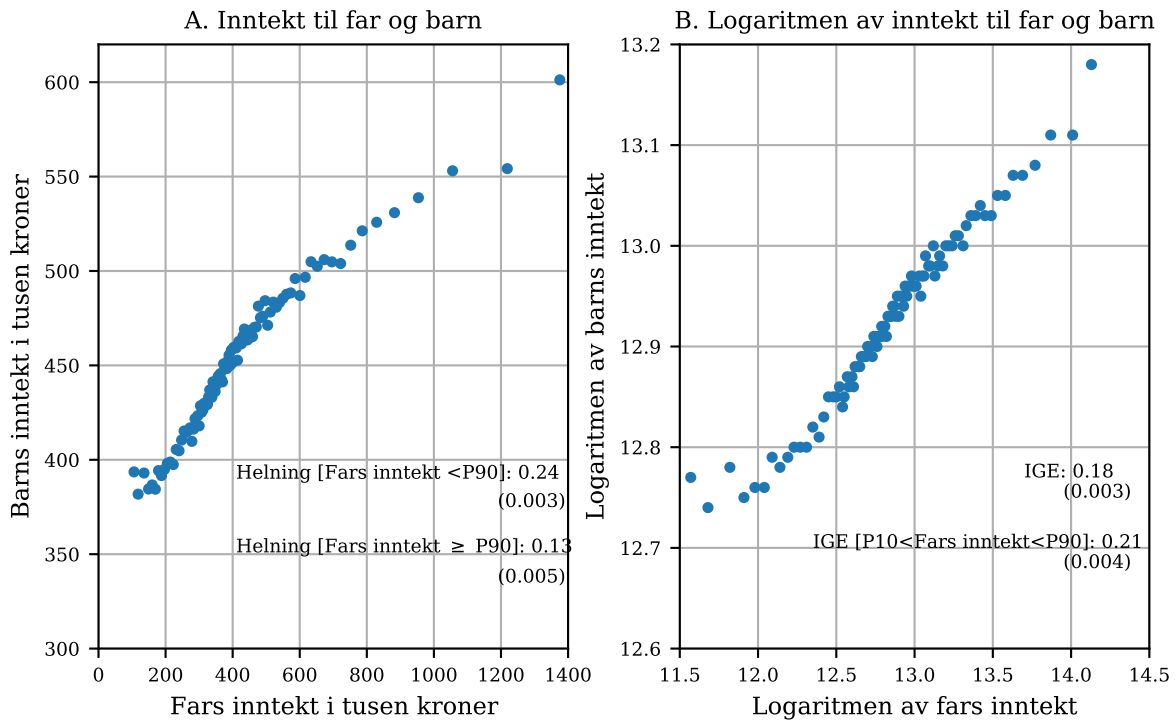
Det enkleste målet er å se direkte på sammenhengen mellom inntekt til foreldre og den gjennomsnittlige inntekten barn som vokser opp med foreldre med den gitte inntekten oppnår som voksne.¹ Panel A i figur 6.1 er konstruert ved å dele foreldre inn etter prosentiler av inntektsfordelingen i deres kohort og deretter beregne den gjennomsnittlige inntekt til foreldrene og barna for observasjonene i hver prosentil. Det viser en klar positiv sammenheng mellom inntekten til foreldrene og inntekten barna selv oppnår. For barn som vokser opp med foreldre i de nederste 90 prosentilene er en økning i foreldrenes inntekt på 1000kr assosiert med en gjennomsnittlig økning av egen inntekt med 240kr. For barn med foreldre i den øverste desilen den tilsvarende økningen på 130kr. Vi ser at den marginale økningen er avtagende slik at sammenhengen ikke er lineær. Det betyr at det ikke er mulig å beskrive sammenhengen med en enkelt parameter.

Et alternativ er derfor å se på sammenhengen mellom logartimene av inntektene. Panel B i figur 6.1 er konstruert på samme måte som beskrevet over og viser at sammenhengen mellom logartimene av inntekt er mer tilnærmet lineær. Den estimerte helningen på hele utvalget er 0.18. Ved å ekskludere observasjonene i den laveste og høyeste desilen blir helningen 0.21. Denne estimerte helningskoeffisienten blir betegnet som den intergenera-

¹Jeg bruker inntekt til far som mål på inntekt til foreldrene. Konseptuelt er vi interessert i sammenhengen mellom inntekt til foreldre og deres barn, og vil derfor betegne det som foreldrenes inntekt selv om jeg kun måler inntekt til far.

sjonelle inntektselastisiteten (IGE) og kan tolkes som den forventede prosentvise økningen i barns inntekt hvis inntekten til foreldrene er én prosent høyere. Høyere IGE medfører lavere intergenerasjonell inntektsmobilitet siden foreldrenes inntekt da har større betydning for den gjennomsnittlige inntekten barna selv oppnår som voksne. Den estimerte

Figur 6.1: Forhold mellom inntekt til foreldre og deres barn



Noter: Observasjonene er delt inn i 100 like store grupper etter plassering på fars inntektsfordeling (prosentiler). For hver gruppe er gjennomsnittet av barnas inntekt plottet mot gjennomsnittet av fars inntekt. Det gir et såkalt 'binned scatterplot' med 100 punkter. Den 90. prosentilen av fars inntekt tilsvarer en inntekt på 695.000 kr. Med logaritmen av inntekt tilsvarer den 10. og 90. prosentilen henholdsvis 12.23 og 13.45.

elastisiteten er litt høyere enn i Norberg-Schulz og Østhus (2018). Det kan delvis forklares med at inntekten til barna måles senere slik at det gir et bedre mål på gjennomsnittlig livsløpsinntekt.²

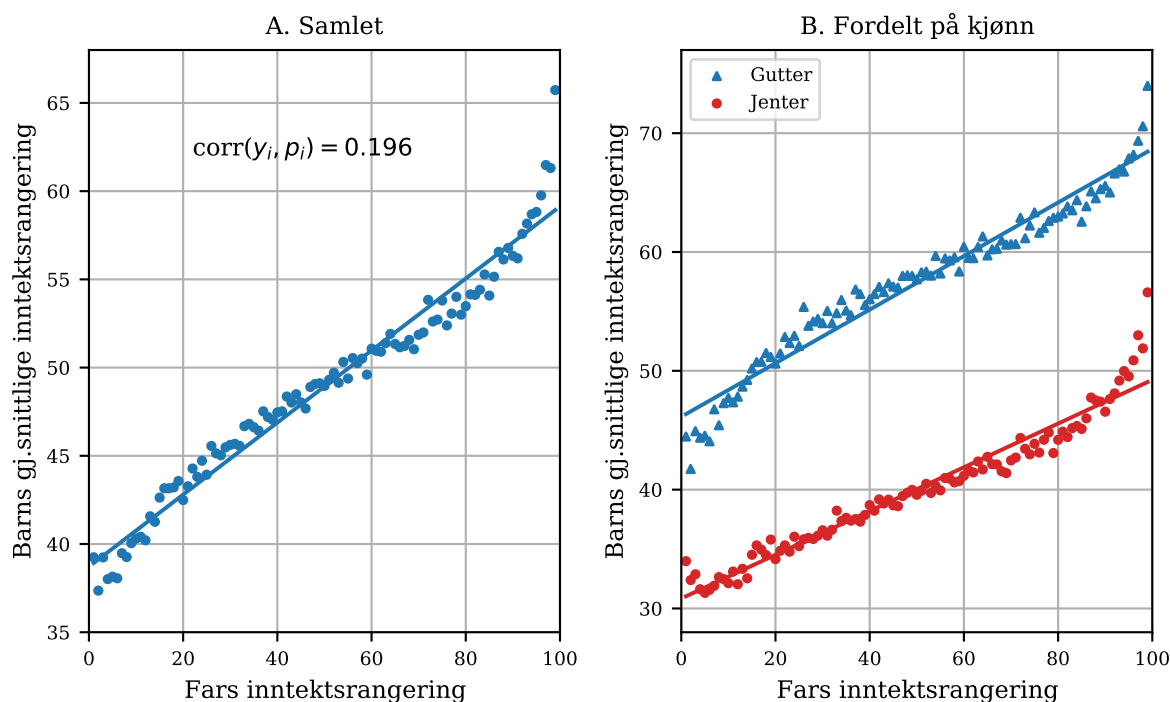
6.2 Inntektsrangering

En alternativ tilnærming er å se på sammenhengen mellom rangeringene til foreldrene og deres barn på sine respektive inntektsfordelinger. Panel A i figur 6.2 er igjen konstruert på samme måte som figurene over, bortsett fra at den for foreldre i hver prosentil nå viser den

²I appendikset bruker jeg pensjonsgivende inntekt som er tilgjengelig fra 1967 til å undersøke effekten på estimatene av IGE ved å måle på inntekt til far og barn i ulike aldre. Det viser at estimatene på IGE øker betydelig når inntekt til barna blir målt i alderen 30-32 år i stedet for 28-30 år, som er målet Norberg-Schulz og Østhus (2018) bruker på de yngste barna i sitt kjerneutvalg.

gjennomsnittlige prosentilen til barna som vokser opp med foreldre i den gitte prosentilen. Den estimerte helningen er 0.2 som medfører at barn som vokser opp hos foreldre som har en inntekt som er 10 prosentiler høyere i inntektsfordelingen til sin kohort i gjennomsnitt havner 2 prosentiler høyere i inntektsfordelingen til egen kohort. Helningskoeffisienten β er et mål relativ inntektsmobilitet og hvis sammenhengen er lineær er $\beta \cdot 100 = \bar{y}_{100} - \bar{y}_1$. Det gir altså et mål på differansen mellom den gjennomsnittlige prosentilen til barn som vokser opp med foreldre med inntekt i den øverste og nederste prosentilen. Panel B i figur 6.2 viser

Figur 6.2: Forhold mellom inntektsrangering til far og barn



Noter: Observasjonene er delt inn i 100 like store grupper etter rangering på fars inntektsfordeling (prosentiler). For hver gruppe er gjennomsnittet av barnas prosentil på egen inntektsfordeling plottet mot fars prosentil. Regresjonslinjene er estimert på de grupperte dataene. Dette kan avvike fra regresjonen på individdata, blant annet fordi fedre i ulike prosentiler i gjennomsnitt har ulik antall barn.

sammenhengen separat for gutter og jenter. Denne figuren viser for det første at gutter har høyere absolutt mobilitet enn jenter for alle prosentiler av foreldrenes inntektsfordeling. Gutter som vokser opp med foreldre i en gitt prosentil vil i gjennomsnitt havne i en høyere prosentil av inntektsfordelingen en jenter som vokser hos foreldre med tilsvarende inntekt. På en annen side har gutter lavere relativ mobilitet i betydningen at helningen er brattere slik at foreldrenes plassering i inntektsfordelingen har større betydning for egen plassering. Forskjellen i helningen er spesielt tydelig for barn med foreldre i de to nederste desilene av inntektsfordelingen.

Regresjonslinjen gir et omtrentlig sammendragsmål på de gjennomsnittlige utfallene til barna som vokser opp med foreldre i hver prosentil, men vi ser at sammenhengen ikke

Tabell 6.1: Nasjonal overgangsmatrise

		Fars kvintil				
		1	2	3	4	5
Barns kvintil	1	29.26%	21.91%	19.12%	17.26%	15.30%
	2	23.64%	23.31%	20.58%	18.55%	15.31%
	3	19.07%	21.46%	21.51%	20.43%	17.43%
	4	16.13%	19.12%	20.87%	21.75%	20.93%
	5	11.90%	14.20%	17.92%	22.01%	31.03%

Noter: Hver celle angir andelen av barna med far med inntekt i kvintilet gitt av kolonnen som selv har inntekt i kvintilene gitt av rekkene.

er helt lineær. For de nederste prosentilene ligger gjennomsnittene under regresjonslinjen og for de øverste prosentilene ligger den over. Vi ser også at sammenhengen er flatere i midten av inntektsfordeling. Det kan tolkes som at sammenhengen mellom plasseringen i inntektsfordelingen til foreldre og barn er sterkere på toppen og bunnen av inntektsfordelingen.

Overgangsmatriser et alternativt mål som er mer informativt om plasseringen på inntektsfordelingen til barn som vokser opp med foreldre i gitte kvantiler. I tabell 6.1 er inntektsfordelingen partisjonert i kvintiler. For hver kvintil av foreldrenes inntektsfordeling viser det sannsynligheten for at barna som vokser opp hos foreldre med inntekt i den gitte kvintilen havner i de ulike kvintilene av inntektsfordelingen i egen kohort. Hvis det var fullstendig mobilitet slik at foreldrenes inntekt ikke har betydning for inntekten barna oppnår som voksne ville det vært 20% i hver celle. Vi ser at barn som vokser opp med foreldre med inntekt i den nederste kvintilen har rundt 29% sannsynlighet for å selv ha inntekt i nederste kvintil av inntektsfordelingen i egen kohort. Sannsynligheten for å gå gjøre klassereise fra nederste til øverste kvintil er på rundt 12%. I andre enden av inntektsfordelingen ser vi at barn fra øverste kvintil har rundt 31% sannsynlighet for å selv havne i øverste kvintil og rundt 15% sannsynlighet for å havne i den nederste. Vi ser at også at sannsynlighetene for å havne i de ulike kvintilene er omtrent er ganske jevnt fordelt for barn som vokser opp med foreldre i 3. og 4. kvintil som tyder på at det høy grad av mobilitet i midten av inntektsfordelingen.

Kapittel 7

Geografisk variasjon i inntektsmobilitet

Formålet med denne analysen er å kartlegge geografisk variasjon i intergenerasjonell inntektsmobilitet innad i landet. Dette gjør jeg ved å dele inn observasjonene i kjerneutvalget etter hvilke sted de vokste opp og deretter se på sammenhengen mellom inntekten til far og inntekten barna oppnår som voksne for hvert av stedene. Det vil da også være mulig å rangere stedene ut fra grad av inntektsmobilitet. Jeg vil først presentere den geografiske inndelingen jeg benytter, og deretter presentere estimat på relativ og absolutt mobilitet for hver av stedene. Til slutt vil jeg diskutere forklaringer på de observerte forskjellene i absolutt mobilitet, og drøfte i hvilken grad det er ulike mekanismer som forklarer variasjon i absolutt mobilitet i Norge og i USA.

7.1 Geografiske enheter

Jeg begynner med å definere oppvekstkommunen som den kommunen barna hadde folkeregistrert adresse da de var 16 år. Siden dette bare angir bosted ved den gitte alderen kan det avvike fra kommunen barnet vokste og det vil også kunne avvike fra bostedskommune når jeg måler inntekten deres som voksen. Barna i kjerneutvalget er 16 år mellom 1988-1996. Jeg velger å omkode kommuneinndelingen til inndelingen som var gjeldende i 2017.

Kommuner er i utgangspunktet en naturlig geografisk inndeling. Utfordringen er at det i mange kommuner er få observasjoner i kjerneutvalget slik det gir upresise estimat på de ulike målene av inntektsmobilitet. Hvis jeg i likhet med Chetty mfl. (2014) setter en nedre grense på 250 observasjoner for å bli inkludert i utvalget faller over hundre kommuner bort. I analysen vil jeg derfor i stedet benytte økonomiske regioner som den geografiske enheten. Dette er en geografisk inndeling på nivået mellom kommune og fylke. Inndelingen bygger på kriterier som handel og felles arbeidsmarked, og det er 89 ulike regioner i alt.¹

¹Inndelingen skal tilsvare det regionale nivået som EU har definert som sin NUTS4-inndeling. En konsekvens av dette er at de økonomiske regionene ikke kan krysse fylkesgrensene.

7.2 Relativ mobilitet

Jeg bruker rangmobilitet som ble beskrevet i seksjon 4.1 for å beskrive inntektsmobiliteten for hvert av stedene. For hver økonomisk region c estimerer jeg parametrene i

$$y_i = \alpha_c + \psi_c p_i + \epsilon_i \quad (7.1)$$

med minste kvadrats metode. Merk at y_i og p_i er prosentiler til henholdsvis barna og foreldrene i de nasjonale inntektsfordelingene til hver av kohortene. Siden sammenhengen er omtrent lineær kan regresjonslinjene brukes som mål på forventet rangering i den nasjonale inntektsfordelingen til barn som vokser i den økonomiske regionen c med foreldre i prosentil p ,

$$\bar{y}_{pc} = E[y|p, c] = \alpha_c + \psi_c p. \quad (7.2)$$

Estimatet på ψ_c kan brukes som mål på den relative mobiliten på de ulike stedene fordi det angir forskjellen i forventet rangering på den nasjonale inntektsfordelingen for barn som vokser opp på toppen og bunnen av inntektsfordelingen,

$$\bar{y}_{100,c} - \bar{y}_{1,c} = 100 \cdot \psi_c. \quad (7.3)$$

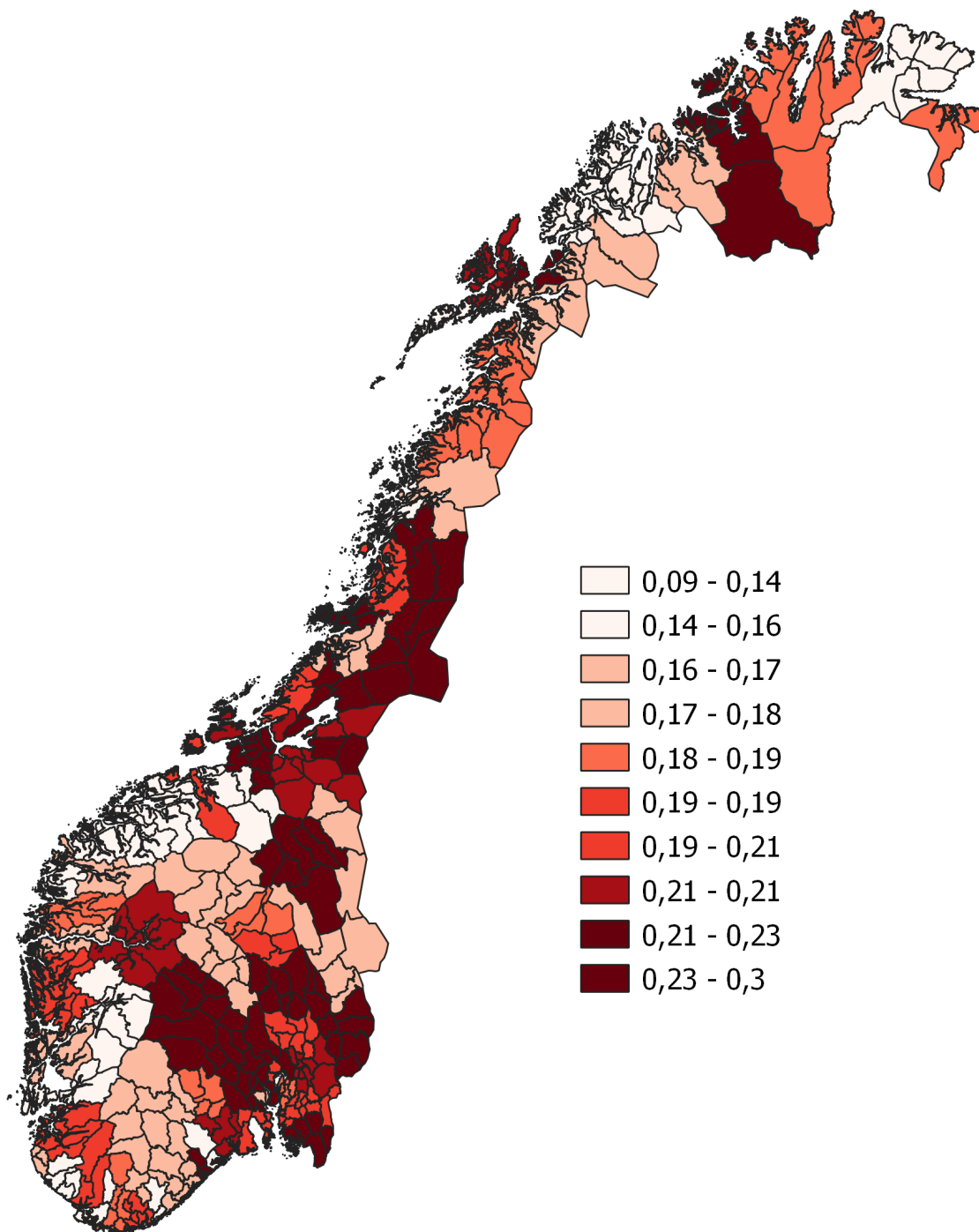
Høyere ψ_c medfører da lavere relativ mobilitet siden det er større differanse i den forventede plasseringen til barn som vokser opp på toppen og bunnen av inntektsfordelingen på det stedet. Kartet i figur 7.1 viser estimatene på ψ_c i de ulike økonomiske regionene. Regionene har blitt rangert etter relativ mobilitet og deretter delt inn i 10 like store grupper, der lysere farger medfører lavere ψ_c og dermed høyere relativ mobilitet. Vi ser at for regionene gruppen med høyest relativ mobilitet er forskjellen i forventet rangering for barna som har vokst opp med foreldre på toppen og bunnen av inntektsfordelingen 9-14 prosentiler. For regionene med lavest relativ mobilitet er den tilsvarende forskjellen mellom 23-30 prosentiler.

7.3 Absolutt mobilitet

De estimerte regresjonslinjene kan også brukes til å rangere de økonomiske regionene etter absolutt mobilitet. Vi kan si at sted j har høyere absolutt mobilitet enn sted k for barn med foreldre i prosentil p hvis

$$\alpha_j + \psi_j p > \alpha_k + \psi_k p, \quad (7.4)$$

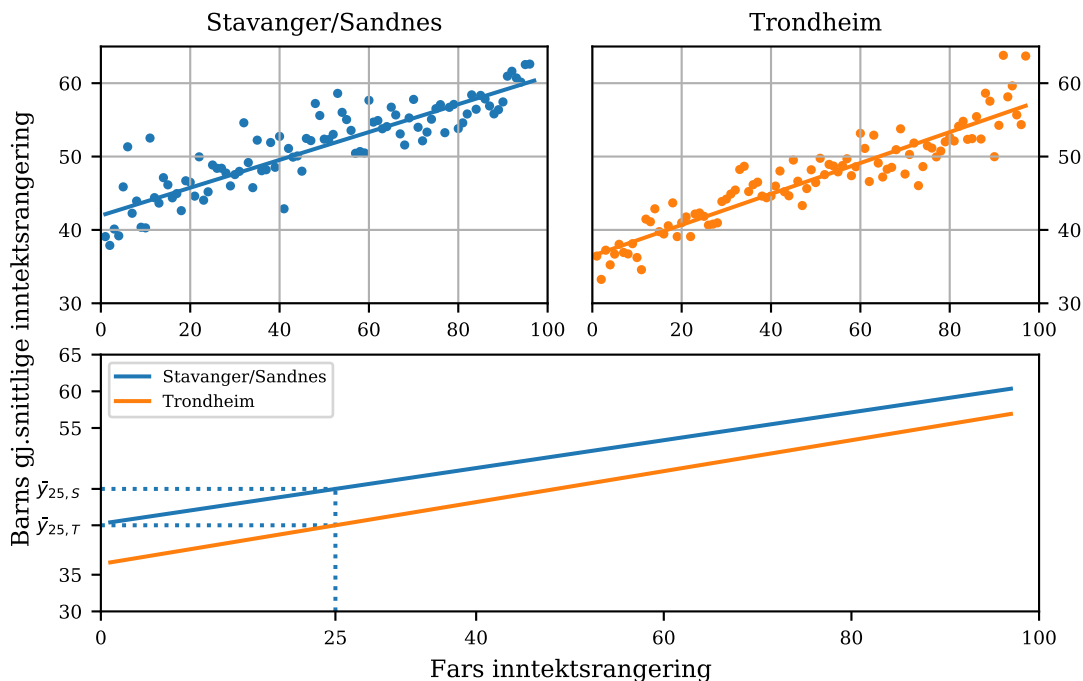
Figur 7.1: Relativ mobilitet



Noter: Kartet viser estimatene på relativ mobilitet ($\hat{\psi}_c$) fra ligning (7.2) i de ulike økonomiske regionene. Estimatene er delt i 10 like store grupper ut fra antall kommuner. Lavere estimat vises med lysere farge og medfører høyere relativ mobilitet. Estimatene er basert på kjerneutvalget og barna er plassert i den økonomiske regionen de hadde folkeregistrert adresse da de var 16 år.

fordi barna på sted j i gjennomsnitt havner høyere opp i den nasjonale inntektsfordelingen enn barn på sted k som vokser opp hos foreldre med tilsvarende inntekt. Hvilken regresjonslinje som ligger høyest kan avhenge av hvilken verdi av p vi betrakter. For å få et entydig mål på absolutt mobilitet som kan brukes til å rangere de ulike økonomiske regionene velger jeg i likhet med Chetty mfl. (2014) å se på verdien av regresjonslinjen i den 25. prosentilen. Dette valget kan motiveres ved at det tilsvarer det gjennomsnittlige utfallet til barna som vokser opp med foreldre i nedre halvdel av inntektsfordelingen. Fremgangsmåten er illustrert i figur 7.2 med de økonomiske regionene Stavanger/Sandnes

Figur 7.2: Absolutt mobilitet i Stavanger/Sandnes og Trondheim



Note: Figuren viser de estimerte regresjonslinjene $y_{pc} = \alpha_c + \psi_c p$ for de økonomiske regionene Stavanger/Sandnes og Trondheim, samt $\bar{y}_{25,c} = \alpha_c + \psi_c \cdot 25$ for begge. Det viser også et såkalt *binned scatterplot* med gjennomsnittlig prosentil til barn med foreldre inntekt i hver prosentil for hver av regionene. Dette er kun for å illustrere i hvilken grad sammenhengen faktisk er lineær. Regresjonslinjene er estimert på individdata. Stavanger/Sandnes består av kommunene Stavanger, Sandnes, Gjesdal, Sola, Randaberg, Forsand, Strand, Hjelmeland, Finnøy, Rennesøy og Kvitsøy. Trondheim består av kommunene Trondheim, Rissa, Midtre Gauldal, Melhus, Skaun, Klæbu, Malvik, Selbu og Tydal.

(S) og Trondheim (T) som eksempler. Vi ser at regresjonslinjen i Stavanger/Sandnes ligger over regresjonslinjen til Trondheim for alle prosentiler av foreldrenes inntekt. Vi kan også se at helningen i linjen til Trondheim er litt brattere som medfører lavere relativ mobilitet. Regresjonslinjene kan brukes til å finne

$$E[y|p = 25, c = S] = \bar{y}_{25,S} = \alpha_S + \psi_S p \quad (7.5)$$

og tilsvarende for Trondheim. Merk at regresjonene er gjort på individdata og at spredningsplottet med gjennomsnittsverdi for hver prosentil er inkludert for å illustrere at sammenhengen er omtrent lineær også innad i regionene.

Jeg har estimert $\bar{y}_{25,c}$ for alle de 89 økonomiske regionene og resultatet vises i kartet i figur 7.3. Regionene er rangert etter absolutt mobilitet og delt inn i 10 like store grupper, der lysere farger medfører høyere absolutt mobilitet.² Fra kartet kan vi se at det er høyest absolutt mobilitet på Vestlandet. Faktisk er de 12 økonomiske regionene med høyest absolutt mobilitet alle i vestlandsfylkene (jmf. Tabell A.1). I gruppen med lavest absolutt mobilitet havner barna med foreldre i 25. prosentil i gjennomsnitt under 42. prosentil av inntektsfordelingen. Dette tilsvarer en inntekt rett i overkant av 400 000 kroner. I regionene med høyest mobilitet havner barna som vokser opp hos foreldre med tilsvarende inntekt i gjennomsnitt over 47. prosentil av inntektsfordelingen som tilsvarer en inntekt på rundt 420 000 kroner. Forskjellen på regionen med høyest og lavest absolutt mobilitet tilsvarer en forskjell i inntekt på 431 000 kroner og 393 000 kroner.

De geografiske forskjellene i absolutt mobilitet er mindre enn i USA. Chetty mfl. (2014) bruker samme mål på absolutt mobilitet, rangerer pendlerregionene og deler de inn i 10 like store grupper. De finner at barn som vokser opp med foreldre i 25. prosentil i gjennomsnitt havner under den 37. prosentil i gruppen av pendlerregioner med lavest mobilitet. For gruppen av pendlerregionene med høyest mobilitet er det gjennomsnittlige utfallet over den 52. prosentilen. Den samlede spredningen i gjennomsnittlig utfall fra pendlerregion med lavest til høyest absolutt mobilitet er fra 26. til 65. prosentil. Siden inntektsforskjellene er større, blir også den korresponderende spredningen i inntekt større med en forskjell på \$55 000 og \$20 000 i pendlerregionene med høyest og lavest absolutt mobilitet.

Resultatene er i tråd med Norberg-Schulz og Østhus (2018), som også finner høyest absolutt mobilitet på Vestlandet. De setter dette i sammenheng at landsdelen har hatt høy inntektsvekst i perioden barna vokser opp, blant annet som følge petroleumsnæringen er geografisk konsentrert på vestlandet. De peker også på at kun 20 prosent av barna i utvalget som vokser opp i Hordaland og Rogaland har flyttet ut av hjemfylkene når de måler inntekten deres som voksne, slik at det rimelig at inntektsveksten løfter barna oppover på den nasjonale inntektsfordelingen. Dette er til dels i kontrast med Chetty mfl. (2014) som finner at inntektsmobilitet på ulike steder i USA i liten grad er korrelert med egenskaper ved det lokale arbeidsmarkedet som sysselsetting og lønnsvekst. De finner derimot at høy absolutt mobilitet er korrelert med (1) lite segregering, (2) lav inntektsulikhet, (3) andel husholdninger med gifte foreldre, (4) mål på sosial kapital som for eksempel valgdeltagelse og (5) skolekvalitet.

²De er delt inn i 10 like store grupper etter antall kommuner. Ettersom estimatene er beregnet for økonomiske regioner hadde det vært mer naturlig å dele inn i like store grupper etter antall regioner, men det var ikke trivielt å omkode dette manuelt i QGIS.

For å illustrere hvorfor egenskaper ved arbeidsmarkedet har lite forklaringskraft bruker de Charlotte i North Carolina som et eksempel. Av de femti største *metropolitan areas* har Charlotte lavest absolutt mobilitet for barn som vokser opp med foreldre i nedre halvdel av inntektsfordelingen, og kun 4.4 prosent av barna av barna med foreldre i nederste kvintil havner selv i øverste kvintil. Dette er på tross av høy inntektsvekst og en voksende finanssektor. De bruker dette eksempelet til å argumente for at det er lite korrelasjon mellom lønnsvekst og absolutt mobilitet ettersom de *gode jobbene* tilfaller innflyttere fra andre deler av landet dersom lokale arbeidstagere mangler nødvendige kvalifikasjoner. Barna som vokser opp i nedre halvdel av inntektsfordelingen er i stor grad ekskludert fra de godt betalte jobbene tilknyttet den lokale finanssektoren og dermed har ikke inntektsveksten medført økt absolutt mobilitet.

De konkluderer med at de geografiske forskjellene i absolutt mobilitet i stor grad kan forklares med forskjeller i oppvekstmiljø som påvirker i hvilken grad barna tilegner seg kvalifikasjoner og personlige egenskaper som påvirker deres lønn som voksne. Denne konklusjonen blir underbygget av studien til Chetty og Hendren (2018) som finner at barn som flytter til steder med høyere mobilitet tidligere i oppveksten oppnår større forbedring i egne utfall og at det er liten effekt av å flytte som ung voksen.

De ulike konklusjonene kan skyldes forskjellige egenskaper ved arbeidsmarkedene i Norge og USA. Arbeidsmarkedet i USA er i større grad todelt, med store lønnsforskjeller mellom personer med og uten universitetsutdanning (Corak 2013). De godt betalte jobbene i blant annet finans- og teknologisektoren er i stor grad knyttet til utdanning. I kontrast til dette er det en flatere lønnsstruktur i Norge og yrker som ikke krever utdanning har relativt bedre lønsvilkår. Oljesektoren på Vestlandet har spesielt gitt stor tilgang til godt betalte jobber som ikke krever universitetsutdanning. Mer generelt så virker det rimelig at det er sterkere sammenheng mellom geografiske forskjeller i inntektsvekst og absolutt mobilitet i Norge, siden det i mindre grad er slik at personer uten universitetsutdanning blir ekskludert fra de godt betalte jobbene og inntektsveksten dermed i større grad løfter alle båter.

På en annen side virker det rimelig at forskjeller i oppvekstmiljø kan forklare noe av variasjon i absolutt mobilitet også i Norge. Norberg-Schulz og Østhus (2018) finner geografisk variasjon i absolutt mobilitet også mellom ulike bydeler i Oslo, Bergen, Stavanger og Trondheim, som i mindre grad kan forklares med forskjeller i egenskaper ved lokale arbeidsmarked. Forskjellene mellom bydelene kan derimot også skyldes seleksjonseffekt ved at familiene som bosetter seg de ulike stedene er systematisk forskjellige betinget av inntekt (jmf. diskusjon i kapittel 3). For å estimere en kausal effekt av oppvekstmiljø er det nødvendig å isolere eksogen variasjon i antall år barn blir eksponert for oppvekstmiljø på ulike steder.

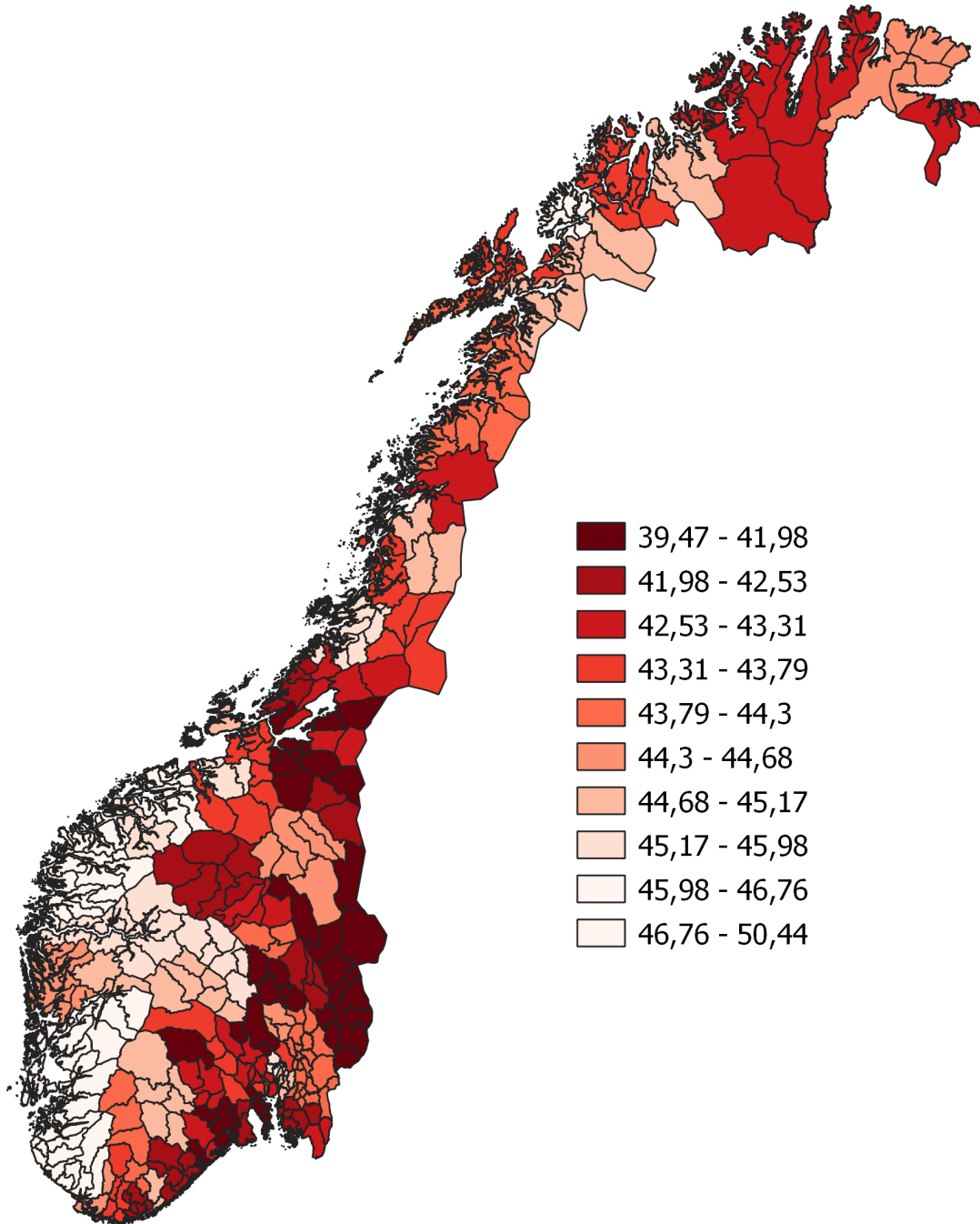
Markussen og Røed (2018) bruker norske data til å estimere kausal effekt av nabolags sosioøkonomisk status på barns langsiktige utfall. De begynner (noe forenklet) ved

å karakterisere nabolag ut fra den gjennomsnittlige inntektsrangeringen til personene i alderen 30-60 år i hvert nabolag. Deretter utnytter de at søsken i familier som flytter mellom ulike nabolag får ulik eksponering for nabolag med ulik sosioøkonomisk status. De finner at såkalte middelklassenabolag i midten av inntektsfordelingen gir størst forbedring i barnas utfall og at rike nabolag gir nesten like dårlige utfall som nabolag i bunnen av inntektsrangeringen.

Den kausale effekten av nabolag på barns utfall betegnes som nabolagseffekter. Nabolag kan påvirke utfall til barna gjennom flere ulike kanaler, blant annet egenskaper ved miljøet (luftkvalitet, støy, tilgang til rekreasjonsområder), offentlige tjenester (barnehager, skoler, helsetjenester) og påvirkning gjennom sosialt miljø (venner, klassekamerater, rollemodeller). Markussen og Røed argumenter for at i velferdsstater som Norge med mål om å utjevne forskjeller og gi lik tilgang på offentlige tjenester vil mekanismen i hovedsak gå gjennom sosiale interaksjonseffekter. I USA er det større forskjeller mellom fattige og rike nabolag langs de andre dimensjonene, for eksempel ved at offentlige skoler er finansiert gjennom lokale skatter og at skoler i rikere nabolag dermed ofte bruker mer ressurser per elev. I Norge er det i praksis motsatt.

Jeg skal nå undersøke om oppvekstmiljø kan forklare noe av variasjonen i absolutt mobilitet ved å se om barn som flytter til kommuner med høyere absolutt mobilitet tidlig i oppveksten får bedre utfall enn de som flytter senere. Kommuneinndeling er en grovere geografisk inndeling enn nabolag, og det kan være til dels betydelig variasjon i oppvekstmiljø innad i hver kommune. Variasjonen i nabolagseffekter i denne grovere geografiske inndelingen vil uansett gi en nedre begrensning på den samlede variasjonen som også inkluderer variasjon innad i hver geografisk enhet. I den grad forskjeller i absolutt mobilitet kan forklares med ulik eksponering for oppvekstmiljø i ulike kommuner kan dette tolkes som nabolagseffekter.

Figur 7.3: Absolutt mobilitet



Noter: Kartet viser estimatene på \bar{y}_{25} fra ligning (7.5) i de ulike økonomiske regionene. Estimatenes er delt i 10 like store grupper ut fra antall kommuner. Høyere estimat vises med lysere farge og medfører høyere absolutt mobilitet. Estimatenes er basert på kjerneutvalget og barna er plassert i den økonomiske regionen de hadde folkeregistrert adresse da de var 16 år.

Kapittel 8

Eksponeringseffekt av oppvekstmiljø

Vi har sett at det er geografiske forskjeller i absolutt mobilitet innad i Norge og at det er ulike mulige forklaringer på disse observerte forskjellene. Jeg skal nå analysere et kvasieksperiment som utnytter variasjon i barns alder når de flytter til å vurdere relativ betydning av de ulike forklaringene. Dette forskningsdesignet ble først brukt av Chetty og Hendren (2018) som finner at den geografiske variasjonen i absolutt mobilitet i USA i stor grad skyldes forskjeller i oppvekstmiljø og at forskjeller i egenskaper ved arbeidsmarked har lite betydning. I motsetning til deres studie observerer jeg ikke barn som flytter etter de er 18 år. Det er derfor ikke mulig å trekke like sterke konklusjoner om betydning av forskjeller i egenskaper ved arbeidsmarked, siden jeg ikke observerer barn som flytter like før de går inn i arbeidsmarkedet. Jeg fokuserer i stedet på om det er mulig å avdekke kausale nabolageffekter. Forskningsdesignet er beskrevet i seksjon 4.2-4.4, men jeg vil gi en kort, ikke-teknisk sammendrag av det her. Deretter presenterer jeg resultatene og forsøker å vurdere troverdigheten til den identifiserende antagelsen.

8.1 Forskningsdesign

Opplegget går ut på å undersøke om barn som flytter til *bedre steder* også selv oppnår bedre utfall. Begrepet *bedre sted* brukes i betydningen av at barna som tilbringer hele oppveksten der i gjennomsnitt oppnår bedre utfall, betinget av inntekten til foreldrene.¹ Det er noen steder som har spesielt bra utfall for barn som vokser opp med foreldre i nedre halvdel av inntektsfordelingen, og andre steder utfallene spesielt bra for barn med foreldre i toppen av inntektsfordelingen. Hva som utgjør et bedre sted avhenger derfor av foreldrenes plassering på inntektsfordelingen. Mer spesifikt vil vi undersøke om barn som flytter til steder der de permanente innbyggerne som vokser opp hos foreldre med tilsvarende inntekt i gjennomsnitt havner høyere i den nasjonale inntektsfordelingen også selv oppnår høyere inntekt. Dette gjør vi ved å se på sammenheng mellom variasjon i barnas

¹Barna som tilbringer hele oppveksten på et sted vil heretter betegnes som permanente innbyggere.

utfall og kvaliteten på stedet barna flyttet til, betinget av kvaliteten på stedet de flyttet fra.² Denne sammenhengen avdekker ikke nødvendigvis noen kausal effekt siden familier som flytter til bedre steder kan være systematisk forskjellige på måter som påvirker barns inntekt, noe som medfører en seleksjonskjevhet. Forskningsdesignet går ut på å estimere sammenhengen separat for hver alder barna flytter og å anta at seleksjonskjevheten er konstant for de ulike flyttealdrene. Differansene i effekten av å flytte til bedre steder ved ulike aldre vil da identifisere kausale eksponeringseffekter.

8.2 Estimat av eksponeringseffekt

Jeg begynner med å estimere b_m som er den gjennomsnittlige endringen i barns prosentil ved å flytte til et sted der de permanente innbyggerne med foreldre i samme prosentil i gjennomsnitt havner én prosentil høyere i den nasjonale inntektsfordeling. Estimaten fra spesifikasjonen

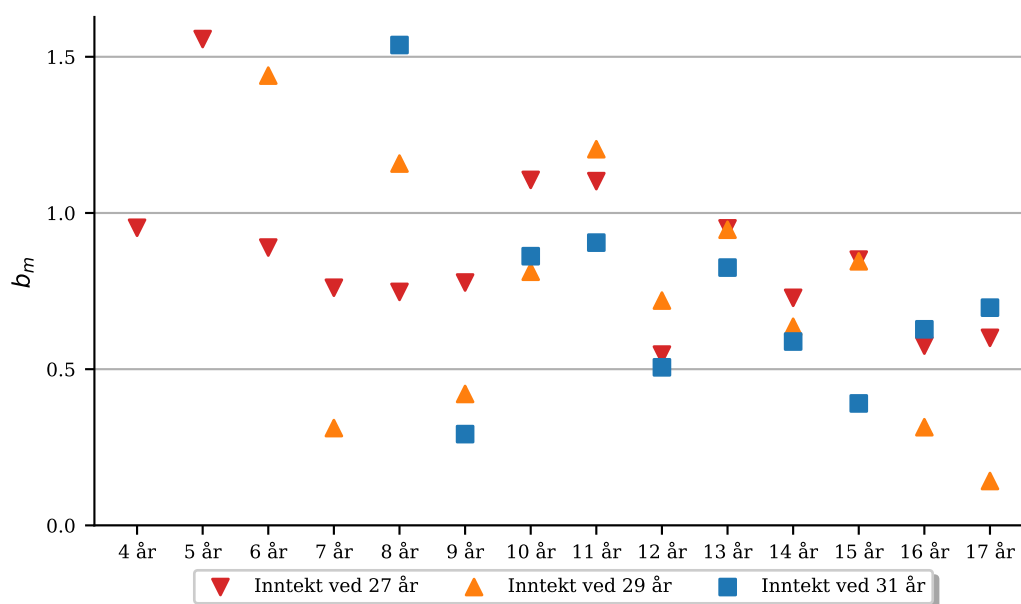
$$y_i = \alpha^1 + \alpha^2 \bar{y}_{po} + \sum_m b_m I(m_i = m) (\zeta_m^1 + \zeta_m^2 p_i) + \sum_m b_m I(m_i = m) \Delta_{pod} + \epsilon_{3i} \quad (8.1)$$

vises i figur i figur 8.1. Disse estimatene har ikke kausal tolkning siden valg av destinasjon er endogent, men gitt at den identifiserende antagelsen er gyldig vil høyere estimat ved tidligere flyttealdre være mål på en positiv kausal eksponeringseffekt. Figuren viser estimatene med inntekt målt i tre ulike aldre. Det er en del spredning i estimatene av b_m ved hver flyttealder m , og spredningen er større for tidlige flyttealdre der det er færre observerte flyttinger. Likevel er det en tendens til estimatene er høyere for tidligere flyttealdre og dette gjelder for alle aldre jeg måler inntekt.

Figur 8.2 viser estimatene separat for utvalgene med inntekt målt i ulike aldre, samt standardavviket til hvert av estimatene. Den viser at estimatene er mer usikre for tidlige flyttealdre der det er færre observerte flyttinger. Det er også mer usikkerhet til estimatene der inntekt er målt senere siden utvalgsstørrelsen er mindre. Usikkerheten til estimatene er stor i forhold til differansen mellom estimatene ved ulike aldre. Dette gjør det vanskelig å påvise statistisk signifikante eksponeringseffekter $\gamma_m = b_m - b_{m+1}$ ved ulike flyttealdre m . Motivert av at Chetty og Hendren (2018) finner at sammenhengen er lineær fram til $m = 23$ og at sammenhengen ser omtrent lineær ut i mine data, vil jeg anta at $\gamma_m = \gamma$ for alle m . Dette betyr at den kausale eksponeringseffekten av å tilbringe ett år ekstra av oppveksten på et sted der de permanente innbyggerne i gjennomsnitt havner én prosentil høyere i inntektsfordelingen ikke avhenger av alder til barna. Eventuelt kan parameteren

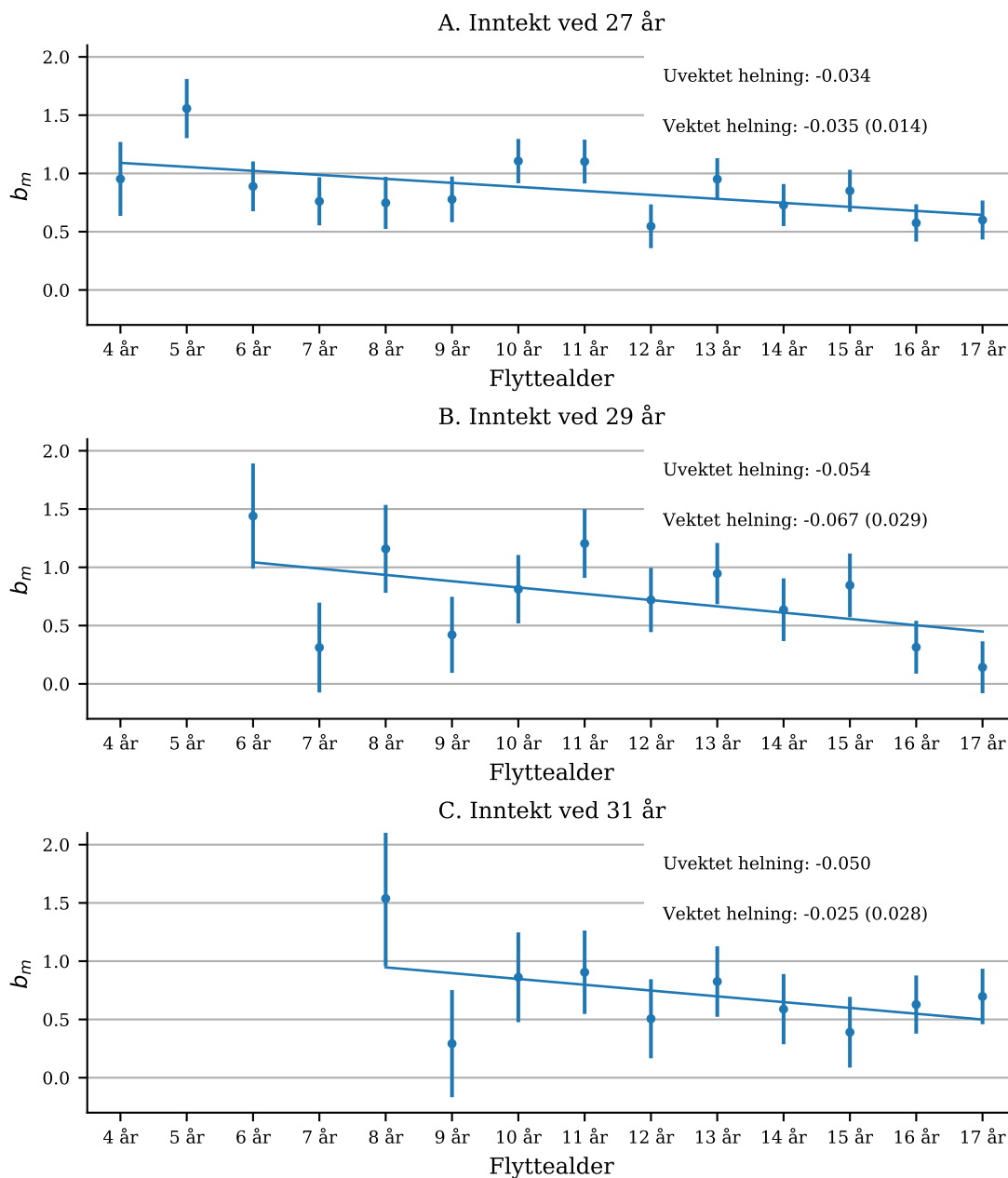
²Kvaliteten brukes her som mål på i hvilken grad stedene er *bra* eller *dårlig*, som avhenger av de gjennomsnittlige utfallene til de permanente innbyggerne som vokser opp med foreldre i samme prosentil.

Figur 8.1: Estimert på b_m (1)



Noter: Figuren viser de estimerte koeffisientene av $\{b_m\}$ fra spesifikasjon (8.1). Utvalget er avgrenset av barna som er født mellom 1977 og det siste året jeg kan måle inntekten til barna i den angitte alderen i 2015. Det vil si 1988, 1986 og 1984 med inntekt ved henholdsvis 27, 29 og 31 år. Utvalget består av barna som flytter én gang mellom 150 største kommunene mellom 1992 og før de blir 18 år. Koeffisientene $\{b_m\}$ kan tolkes som effekt av å flytte i alder m til en kommune der de permanente innbyggerne med foreldre med samme inntekt i gjennomsnitt havner én prosentil høyere i den nasjonale inntektsfordelingen. Merk at jeg estimerer rangmobiliteten til de permanente innbyggerne med inntekt i hver alder. Forskjellene i estimatene med inntekt i ulik alder skyldes dermed både forskjeller i utfall til flytterne og forskjell i estimatene på rangmobilitet til de permanente innbyggerne. I figur A.1 viser jeg estimater med inntekt i ulik alder med samme kohort-avgrensning på utvalget av flyttere.

Figur 8.2: Estimat på b_m (2)



Noter: Figuren viser de samme estimatene som i figur 8.1 og viser standardfeilene til hver av estimatene. Den viser også regresjonslinjene som er estimert med minste kvadrats metode av $\{b_m\}$ på flyttealder m . Den vektete helningen er fra en tilsvarende regresjon der hvert estimat er vektet med antallet barn som flytter i alderen gjennom implementeringen av analytiske vektorer i Stata.

Tabell 8.1: Estimat på gjennomsnittlig eksponeringseffekt

	Inntekt ved alder:		
	27 år (1)	29 år (2)	31 år (3)
Vektet helning	0.035 (0.014)	0.067 (0.029)	0.025 (0.028)
Parametrisk	0.034 (0.013)	0.064 (0.025)	0.027 (0.037)
Søsken	-0.005 (0.034)	0.021 (0.059)	-0.043 (0.116)
Antall observasjoner	27638	19114	13155
Antall søsken	9960	5871	3393

Noter: Standardfeil i parantes. Tabellen presenterer ulike estimat på den den gjennomsnittlige eksponeringseffekten ved å tilbringe ett år lengre på et sted der de permanente innbyggerne i gjennomsnitt havner én prosentil høyere i den nasjonale inntektsfordelingen. Utvalget av flyttere består av barna som er født mellom 1977 og det siste året jeg kan måle inntekt ved angitt alder i 2015 og som flytter én gang mellom de 150 største kommunene mellom 1992 og før året de fyller 18 år. Vektet helning er den negative helningen av en regresjon av estimatene på b_m fra spesifikasjon (8.1) på flyttealder m , vektet med antallet barn som flytter i hver alder. Det parametriske estimatet er fra spesifikasjon (8.2). Søskenestimatet er fra spesifikasjon (8.3), der standardfeilene er justert for antall frihetsgrader med utgangspunkt i at det i gjennomsnitt er to søsken per familie (se appendiks for detaljer).

også tolkes som et gjennomsnitt av eksponeringseffektene for ulike aldre.

Figur 8.2 viser også regresjonslinjer som er estimert med minste kvadrats metode av estimatene b_m på flyttealder m . Dette illustrerer at sammenhengen er avtagende og den omvendte verdien av helningen vil være et estimat på γ . Dette estimatet tar ikke hensyn til at de ulike estimatene av b_m er målt med ulik presisjon. Jeg har derfor også rapportert helningen fra en vektet regresjon, der hver b_m er vektet med antallet barn i utvalget som flytter i alderen m .³ Dette gir estimat av γ på mellom 0.025–0.067 avhengig av når inntekt blir målt. Estimaten er forholdsvis usikre, men med inntekt målt ved 27 og 29 år er de statistisk signifikante på 0.05 nivå. Til sammenligning finner Chetty og Hendren estimat på mellom 0.03-0.045 avhengig av spesifisering og avgrensing av utvalg. Deres estimat er langt mer presise siden de opererer med utvalg på over 700 000 barn som flytter.

En alternativ fremgangsmåte for å estimere γ er å modifisere spesifisjonen (8.1) til å måle effekten av å tilbringe ett år lengre på et sted der de permanente innbyggerne med foreldre med tilsvarende inntekt i gjennomsnitt havner én prosentil høyere i den nasjonale inntektsfordelingen. Siden jeg observerer flyttinger fram til barna er 17 år gir det spesifisjonen

$$y_i = \alpha^1 + \alpha^2 \bar{y}_{po} + \sum_m b_m I(m_i = m) (\zeta_m^1 + \zeta_m^2 p_i) + \gamma(18 - m_i) \Delta_{pod} + \kappa \Delta_{pod} + \epsilon_{3i}. \quad (8.2)$$

Estimatene av γ fra denne spesifisjonen er rapportert som det parametriske estimatet i tabell 8.1. Disse estimatene samsvarer i stor grad med den omvendte helningen av den vektete regresjonen. Estimaten er lavest med inntekt målt ved 31 år, som er målet på inntekt som gir best mål på gjennomsnittlig livsinntekt til barna. Jeg tar derfor utgangspunkt i et konservative estimat av γ på rundt 0.03 som er i tråd med de laveste estimatene Chetty og Hendren finner i deres tilsvarende analyse. Dette medfører en konvergensrate på 3 % per år med eksponering. Med andre ord er det slik at utfallene til innflytterne konvergerer mot de gjennomsnittlige utfallene til de permanente innbyggerne og γ gir mål på konvergensraten. De finner konvergens også langs andre utfall enn inntekt, og konkluderer med forskjeller i utfall mellom barn som vokser opp ulike nabolag i stor grad skyldes kausale stedeffekter i stedet for forskjeller i egenskapene til innbyggerne. De konkluderer også med at sted har i større grad har betydning for absolutt mobilitet gjennom forskjeller i oppvekstmiljø i stedet for forskjeller i egenskap ved arbeidsmarkedene på ulike steder. Siden estimatet eksponeringseffekten er lignende på norske data kan det være rimelig å generalisere disse konklusjonene.

På en annen side er estimatene av b_m gjennomgående høyere på norske data enn i analysen til Chetty og Hendren. Det betyr at det for hver flyttealder m er større effekt

³Det er estimert med implementeringen av analytiske vekter i Stata.

av å flytte til et bedre sted. Denne effekten er sammensatt av en kausal effekt av flytting. De høyere estimatene kan derfor både skyldes at det er større kausal effekt av flytting i Norge, for eksempel ved at forskjeller i egenskaper ved arbeidsmarked på ulike steder har større betydning, eller at det er større seleksjonsskjevhet. Chetty og Hendren observerer flyttinger etter de måler inntekt som gjør det mulig å identifisere seleksjonseffekten som også gjør det mulig å dekomponere b_m gitt den identifiserende antagelsen om at seleksjonseffekten er konstant for ulike flyttealdrer. På grunn av begrensinger ved datagrunnlaget mitt er det ikke mulig å gjennomføre denne dekomponeringen, så jeg kan ikke trekke like sterke konklusjoner om betydningen av forskjeller i arbeidsmarkeder på ulike steder på de observerte forskjellene i absolutt mobilitet.

8.3 Validering av forskningsdesignet

Konklusjonene over forutsetter at den identifiserende antagelsen om at seleksjonseffekten ikke avhenger av alderen barna flytter er gyldig. Den er ikke gyldig dersom familiene som flytter til bedre steder når barna er tidlig oppveksten er systematisk forskjellig fra familiene som flytter når barna er eldre på måter som påvirker utfallene til barna. Det kan for eksempel være slik at foreldrene som flytter til bedre steder når barna er yngre i gjennomsnitt har høyere utdannelse. En annen trussel mot identifikasjonen er at beslutningen om å flytte til et bedre sted er assosiert med andre endringer i oppvekstmiljøet innad i familien, for eksempel ved at foreldre får ny jobb eller endret samlivsstatus. Da kan forskjeller i utfall for barn som flytter i ulik alder delvis skyldes ulik eksponering for andre endringer i oppvekstmiljø enn stedseffekter.

For å formalisere diskusjonen kan vi dekomponere effekten av uobserverte variabler som påvirker barns utfall (θ_i) i ligning (4.6) i to deler: en komponent $\bar{\theta}_i$ som er konstant innad i familien og residual $\tilde{\theta}_i = \theta_i - \bar{\theta}_i$ som kan variere over tid innad i familien. Chetty og Hendren har flere ulike strategier for å sikre at θ_i ikke er korrelert med alder barna flytter til bedre steder, både som følge av korrelasjon med den konstante komponenten og residualen. De finner at estimatene er robuste for de ulike strategiene og dette øker troverdigheten til identifiserende antagelsen.

Jeg skal kontrollere for forskjeller i den konstante komponenten $\bar{\theta}_i$ ved å sammenligne utfall til søsken som flytter i ulik alder. Anta for eksempel at en familie med to barn flytter når barna er m_1 og m_2 år gamle, der $m_1 < m_2$. Den yngre søskenen får da lengre eksponering for oppvekstmiljøet på det nye stedet. Den kausale eksponeringseffekten γ er da identifisert av sammenhengen mellom forskjellene i søskenens utfall, $y_1 - y_2$, og interaksjonen mellom forskjellen i flyttealder og forskjellen i utfallene til permanente innbyggerne på opprinnelse og destinasjon, $(m_1 - m_2)\Delta_{pod}$.

I utgangspunktet kunne jeg implementere søskensammenligningene ved å legge til familie fasteffekter på de tidligere spesifikasjonene. På grunn av begrensinger ved microdata.no

er det ikke mulig å lage panel med variabler som ikke har tidsforløp eller konstruere dummyvariabler for hver familie. Jeg kan derimot finne gjennomsnitt av variabler innad i familier og bruke dette til å konstruere variabler som angir avvik fra gjennomsnitt innad i hver familie. La j indikere hvilken familie barnet tilhører og la antall år av oppveksten med eksponering for det nye stedet være $e_{ij} = 18 - m_{ij}$. Jeg kan da estimere eksponeringseffekten med spesifikasjonen

$$y_{ij} - \bar{y}_j = \alpha + \gamma (e_{ij}\Delta_{pod} - \overline{e_j\Delta_{pod}}) + \epsilon_{4ij}. \quad (8.3)$$

Estimatene fra søskensammenligningene er rapportert i tabell 8.1. Disse estimatene er gjennomgående lavere enn estimatene fra hele utvalget, og de er til dels negative. Dette kan tolkes som indikasjon på at den identifiserende antagelsen i mindre grad er oppfylt på norske data ved at familier som flytter til *bedre steder* når barna er yngre systematisk har høyere $\bar{\theta}_j$. Når vi kontrollerer for dette forsvinner den positive eksponeringseffekten. På en annen side er estimatene fra søskensammenligningene enda mer usikre, så det er ikke mulig å trekke noen sterke konklusjoner om dette.

8.4 Sammendrag

Estimatene på γ i tabell 8.1 for hele utvalget av flyttere er i tråd med resultatene i Chetty og Hendren (2018). Gitt at den identifiserende antagelsen holder, viser disse estimatene at utfallene til innflytterne konvergerer mot utfallene til de permanente innbyggerne og at graden av konvergens avhenger positivt av antall år med eksponering. Det er mulig å tolke dette som bevis for at det eksisterer kausale stedseffekter som påvirker utfallene til alle barna som blir eksponert for det; både innflytterne og de permanente innbyggerne. Dette kan igjen settes i sammenheng med litteratur om betydning av oppvekstmikjø i nabolag på barns langsiktige utfall. I motsetning til deres studie er estimatene lavere når jeg kontrollerer for familiebakgrunn ved å sammenligne utfall til søsken, men dette kan skyldes tilfeldigheter siden estimatene er usikre.

Estimatene på $\{b_m\}$ i figur 8.1 og 8.2 viser at barna som flytter til bedre steder også selv oppnår høyere inntekt.⁴ Jamført med diskusjonen i avsnittet over er det større effekt ved å flytte tidligere i oppveksten, men det er også betydelig effekt assosiert ved å flytte til bedre steder sent i oppveksten. Disse estimatene er gjennomgående høyere enn tilsvarende estimater i Chetty og Hendren (2018). I motsetning til differansen av estimatene kan ikke disse gis en kausal tolkning fordi jeg ikke vil anta at familier som flytter til bedre steder ikke er systematisk forskjellig fra familiene som flytter til dårligere steder. De vil derfor være sammensatt av både en behandlingseffekt og en seleksjonseffekt. Chetty og Hendren (2018) argumenterer for at det er mulig å dekomponere effekten fordi de kan observere

⁴Symmetrisk oppnår barn som flytter til dårligere steder lavere inntekt

seleksjonseffekt ved å se på estimatene til barn som flytter etter inntekt blir målt og antar at seleksjonseffekten er konstant for alle flyttealdre. Jeg kan ikke gjøre den samme dekomponeringen, så jeg kan ikke utelukke at de høyere estimatene skyldes en større seleksjonseffekt.

Kapittel 9

Konklusjoner

Norge er blant landene med høyest inntektsmobilitet. Det betyr at sammenlignet med de fleste andre land har inntekten til foreldre mindre betydning for inntekten barna deres oppnår som voksne og det er større bevegelse langs den sosiale rangstigen. På en annen side er det geografisk variasjon i graden av inntektsmobilitet mellom ulike steder i Norge. Vi har sett at barn som vokser opp hos foreldre med tilsvarende inntekt i gjennomsnitt oppnår ulik inntekt som voksne avhengig av hvor de vokser opp, og i likhet med tidligere studier fant vi høyest absolutt mobilitet på Vestlandet. I tillegg til å dokumentere denne geografiske variasjonen har vi undersøkt betydningen av ulike forklaringer på de observerte forskjellene i absolutt mobilitet. Jeg har argumentert for at det grovt sett kan skyldes (1) forskjeller i egenskaper til familier som er bosatt på ulike steder, (2) forskjeller i egenskaper ved lokale arbeidsmarkeder og (3) forskjeller i egenskaper ved oppvekstmiljø. For å undersøke den relative betydningen til de ulike forklaringene har jeg gjennomført et kvasiexperiment som utnytter at barn som flytter i ulike aldre av oppveksten får ulik eksponering for oppvekstmiljø på ulike steder. Dette forskningsdesignet ble først benyttet av Chetty og Hendren (2018) på amerikanske data. I likhet med deres studie finner jeg at utfallene til innflytterne konvergerer mot utfallene til de permanente innbyggerne. Mine estimat av denne konvergensraten er på rundt 3% for hvert år med eksponering, sammenlignet med estimat på 3-4% i deres studie. Dette kan tolkes som bevis for at det er ulik kausal effekt av eksponering for oppvekstmiljø på ulike steder. De konkluderer også med at forskjeller i egenskaper ved arbeidsmarkeder i liten grad kan forklare den geografiske variasjon i absolutt mobilitet. Siden jeg ikke observerer barn som flytter etter at inntekten deres blir målt er det ikke mulig å trekke like sterke konklusjoner om dette.

Forskningsdesignet bygger på den identifiserende antagelsen at seleksjonseffekten er konstant for barn som flytter i ulike aldre. Chetty og Hendren (2018) har flere ulike strategier for å vurdere troverdigheten til denne antagelsen og finner at estimatene er robuste for de ulike strategiene. Jeg gjennomfører én av disse strategiene ved å kontrollere for familiebakgrunn gjennom å sammenligne utfall til søsken som er i ulik alder når familien flytter. Med denne spesifikasjonen forsvinner den positive konvergensraten. Dette kan

tolkes som en indikasjon på at familier som flytter til bedre steder når barna er yngre er systematisk forskjellig fra familiene som flytter når barna er eldre, og at effekten forsvinner når jeg kontrollerer for dette. På en annen side er disse estimatene så upresise at det ikke er mulig å trekke noen sterke konklusjoner.

Mine estimater er langt mer usikre enn tilsvarende estimater i Chetty og Hendren (2018). Dette skyldes delvis at Norge er et mindre land slik at det ikke er mulig å observere like mange barn som flytter under oppveksten. Det skyldes også begrensinger ved tilgangen til data, ved at bosted kun er tilgjengelig fra 1992 og at det kun foreligger data om inntekt fram til 2015. Med nyere mål på inntekt vil det være mulig å observere flere barn som flytter. I fremtiden vil det derfor være mulig å gjennomføre en tilsvarende analyse med større utvalg, oppnå mer presise estimater og trekke sterkere konklusjoner.

Bibliografi

Aasland, Sigrun Gjerløw. *Det trengs en landsby : hvordan familie, skole og nabolag påvirker våre barns fremtid*. nob. Oslo, 2019.

Becker, Gary S. og Tomes, Nigel. “An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility”. eng. I: *Journal of Political Economy* 87.6 (1979), s. 1153–1189.

Becker, Gary S. og Tomes, Nigel. “Human Capital and the Rise and Fall of Families”. eng. I: *Journal of Labor Economics* 4.3, Part 2 (1986), S1–S39.

Bhattacharya, Debopam og Mazumder, Bhashkar. “Nonparametric analysis of intergenerational income mobility with application to the United States”. eng. I: *IDEAS Working Paper Series from RePEc* (2007).

Black, Sandra og Devereux, Paul. “Recent Developments in Intergenerational Mobility”. eng. I: *IDEAS Working Paper Series from RePEc* (2010).

Bratberg, Espen, Davis, Jonathan, Mazumder, Bhashkar, Nybom, Martin, Schnitzlein, Daniel D og Vaage, Kjell. “A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US”. eng. I: *The Scandinavian Journal of Economics* 119.1 (2017), s. 72–101.

Chetty, Raj, Friedman, John, Hendren, Nathaniel, Jones, Maggie og Porter, Sonya. “The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility”. eng. I: *IDEAS Working Paper Series from RePEc* (2018).

Chetty, Raj og Hendren, Nathaniel. “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects *”. I: *The Quarterly Journal of Economics* 133.3 (2018), s. 1107–1162.

Chetty, Raj, Hendren, Nathaniel og Katz, Lawrence F. “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment”. eng. I: *American Economic Review* 106.4 (2016).

Chetty, Raj, Hendren, Nathaniel, Kline, Patrick, Saez, Emmanuel og Turner, Nicholas. “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility †”. eng. I: *American Economic Review* 104.5 (2014), s. 141–147.

-
- Corak, Miles. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility". eng. I: *Journal of Economic Perspectives* 27.3 (2013), s. 79–102.
- Dahl, Molly og Deleire, Thomas. "The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data". I: *Institute for Research on Poverty Discussion Paper No. 1342-08* (jan. 2008).
- Haider, Steven og Solon, Gary. "Life-Cycle Variation in the Association between Current and Lifetime Earnings". eng. I: *The American Economic Review* 96.4 (2006), s. 1308–1320.
- Jääntti, Markus, Bratsberg, Bernt, Røed, Knut, Raaum, Oddbjørn, Naylor, Robin, Österbacka, Eva, Björklund, Anders og Eriksson, Tor. "American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States". eng. I: *IDEAS Working Paper Series from RePEc* (2006).
- Kling, Jeffrey R, Liebman, Jeffrey B og Katz, Lawrence F. "Experimental Analysis of Neighborhood Effects". I: *Econometrica* 75.1 (2007), s. 83–119.
- Ludwig, Jens, Duncan, Greg J, Genetian, Lisa A, Katz, Lawrence F, Kessler, Ronald C, Kling, Jeffrey R og Sanbonmatsu, Lisa. "Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity". eng. I: *American Economic Review* 103.3 (2013), s. 226–231.
- Markussen, Simen og Røed, Knut. *The Golden Middle Class Neighborhood: Trends in Residential Segregation and Consequences for Offspring Outcomes*. IZA Discussion Papers 11684. Institute of Labor Economics (IZA), jul. 2018.
- Mazumder, Bhashkar. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data". I: *Review of Economics and Statistics* 87.2 (2005), s. 235–255.
- Nilsen, Øivind Anti, Vaage, Kjell, Aakvik, Arild og Jacobsen, Karl. "Intergenerational Earnings Mobility Revisited: Estimates Based on Lifetime Earnings*". I: *Scandinavian Journal of Economics* 114.1 (2012), s. 1–23.
- Oreopoulos, Philip. "The Long-Run Consequences of Living in a Poor Neighborhood". eng. I: *The Quarterly Journal of Economics* 118.4 (2003), s. 1533–1575.
- Sampson, Rj, Morenoff, JD og Gannon-Rowley, T. "Assessing neighborhood effects": Social processes and new directions in research". English. I: *Annual Review Of Sociology* 28 (2002), s. 443–478.
- Sharkey, Patrick og Faber, Jacob W. "Where, When, Why, and For Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects". eng. I: *Annual Review of Sociology* 40.1 (), s. 559–579.

Solon, Gary. “A model of intergenerational mobility variation over time and place”. eng. I: *Generational Income Mobility in North America and Europe*. Cambridge University Press, 2004, s. 38–47.

Solon, Gary. “Intergenerational income mobility in the United States”. I: *American Economic Review* 82.3 (1992), s. 393.

Swift, Adam. *Political philosophy : a beginners' guide for students and politicians*. eng. Cambridge, 2014.

Zimmerman, David J. “Regression Toward Mediocrity in Economic Stature”. eng. I: *The American Economic Review* 82.3 (1992), s. 409–429.

Tillegg A

Appendiks

Tabell A.1: Økonomiske regioner

	Region	Fylke	y25	Andel askeladder	n
1	Ulsteinvik	Møre og Romsdal	50.44	22.06%	2876
2	Florø	Sogn og Fjordane	49.06	17.37%	1467
3	Jæren	Rogaland	48.57	17.43%	4490
4	Egersund	Rogaland	48.18	15.74%	2344
5	Molde	Møre og Romsdal	48.09	16.22%	6198
6	Odda	Hordaland	47.70	17.88%	1350
7	Førde	Sogn og Fjordane	47.67	14.68%	2568
8	Nordfjord	Sogn og Fjordane	47.66	14.09%	2971
9	Ørsta/Volda	Møre og Romsdal	46.89	14.84%	1989
10	Haugesund	Rogaland	46.74	17.16%	9292
11	Stavanger/Sandnes	Rogaland	46.70	16.04%	19187
12	Sunnhordland	Hordaland	46.62	15.97%	6191
13	Finnsnes	Troms	46.29	14.57%	2037
14	Kristiansund	Møre og Romsdal	46.26	14.29%	3505
15	Flekkefjord	Vest-Agder	46.13	11.76%	1628
16	Sandnessjøen	Nordland	46.04	11.59%	1586
17	Ålesund	Møre og Romsdal	45.97	15.32%	8156
18	Surnadal	Møre og Romsdal	45.95	15.24%	1018
19	Bærum/Asker	Akershus	45.93	16.56%	10976
20	Høyanger	Sogn og Fjordane	45.92	8.53%	909
21	Sogndal/Årdal	Sogn og Fjordane	45.82	13.59%	2535
22	Rørvik	Nord-Trøndelag	45.78	10.05%	969
23	Namsos	Nord-Trøndelag	45.27	12.94%	2116
24	Valdres	Oppland	45.21	11.13%	1772
25	Voss	Hordaland	45.17	16.12%	1473
26	Hallingdal	Buskerud	45.16	8.73%	1758
27	Andselv	Troms	45.14	9.49%	1468
28	Follo	Akershus	44.95	14.40%	9111
29	Narvik	Nordland	44.91	10.81%	2805
30	Lillesand	Aust-Agder	44.88	13.74%	1404
31	Nord-Troms	Troms	44.82	11.59%	1087

	Region	Fylke	y25	Andel askeladder	n
32	Frøya/Hitra	Sør-Trøndelag	44.78	10.89%	818
33	Mosjøen	Nordland	44.76	8.06%	1551
34	Vest-Telemark	Telemark	44.70	12.17%	1474
35	Bergen	Hordaland	44.66	13.81%	31305
36	Vadsø	Finnmark	44.65	12.02%	1483
37	Lyngdal/Farsund	Vest-Agder	44.61	19.35%	2095
38	Hadeland	Oppland	44.50	11.20%	2629
39	Tynset	Hedmark	44.39	9.72%	1397
40	Bodø	Nordland	44.30	9.98%	7585
41	Ullensaker/Eidsvoll	Akershus	44.28	10.78%	4704
42	Lillehammer	Oppland	44.17	8.43%	3121
43	Setesdal	Aust-Agder	44.13	25.60%	844
44	Lillestrøm	Akershus	44.13	12.14%	14923
45	Lofoten	Nordland	43.94	10.59%	2445
46	Askim/Mysen	Østfold	43.92	12.80%	4402
47	Sande/Svelvik	Vestfold	43.79	18.14%	1399
48	Sunnalsøra	Møre og Romsdal	43.75	9.87%	1024
49	Orkanger	Sør-Trøndelag	43.62	9.44%	2101
50	Mandal	Vest-Agder	43.60	12.63%	2279
51	Vesterålen	Nordland	43.51	10.33%	3075
52	Tromsø	Troms	43.48	11.16%	5859
53	Harstad	Troms	43.44	9.90%	3224
54	Oppdal	Sør-Trøndelag	43.43	7.69%	905
55	Kongsberg	Buskerud	43.43	11.15%	2710
56	Oslo	Oslo	43.41	14.63%	25818
57	Brønnøysund	Nordland	43.39	9.09%	1276
58	Grong	Nord-Trøndelag	43.32	10.53%	553
59	Kragerø	Telemark	43.28	9.52%	1621
60	Steinkjer	Nord-Trøndelag	43.17	9.90%	4142
61	Hammerfest	Finnmark	43.08	9.57%	2294
62	Notodden/Bø	Telemark	43.00	9.35%	2142
63	Kirkenes	Finnmark	42.97	7.60%	862
64	Drammen	Buskerud	42.89	12.22%	13453
65	Mo i Rana	Nordland	42.81	7.14%	3020
66	Midt-Gudbrandsdalen	Oppland	42.77	10.43%	1380
67	Stjørdalshalsen	Nord-Trøndelag	42.76	6.40%	2183
68	Alta	Finnmark	42.67	8.92%	2100
69	Halden	Østfold	42.55	13.09%	2497
70	Røros	Sør-Trøndelag	42.42	6.72%	686
71	Hamar	Hedmark	42.35	8.91%	7850
72	Kristiansand	Vest-Agder	42.33	11.72%	8824
73	Arendal	Aust-Agder	42.28	10.88%	7099
74	Moss	Østfold	42.26	12.01%	4510
75	Brekstad	Sør-Trøndelag	42.24	6.50%	1626
76	Fredrikstad/Sarpsborg	Østfold	42.06	8.45%	11604
77	Nord-Gudbrandsdalen	Oppland	42.03	8.56%	2044
78	Sandefjord/Larvik	Vestfold	41.99	10.85%	8947
79	Kongsvinger	Hedmark	41.98	9.12%	4579

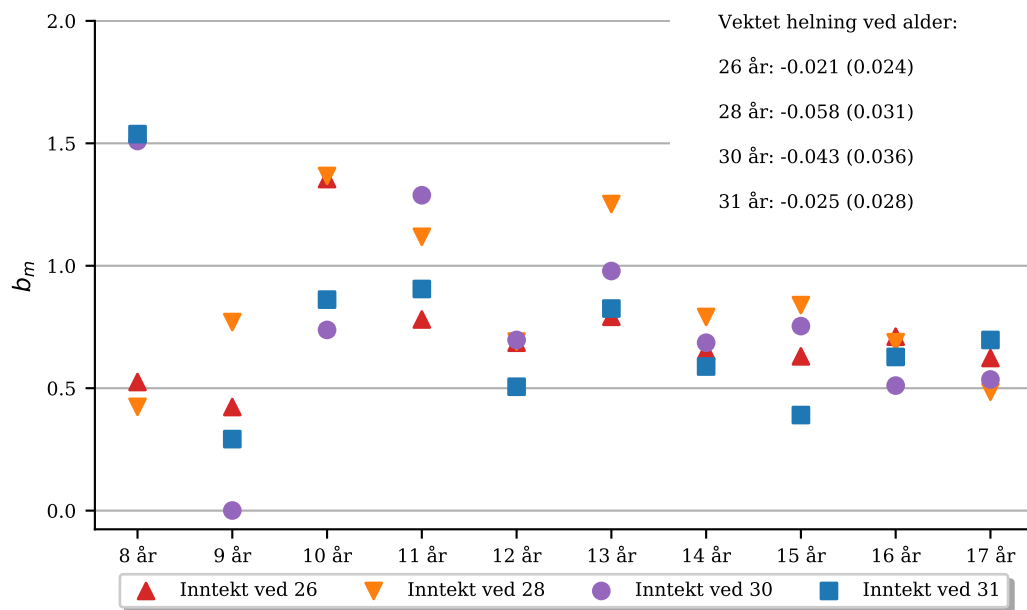
Region	Fylke	y25	Andel askeladder	n
80 Skien/Porsgrunn	Telemark	41.94	8.43%	9961
81 Gjøvik	Oppland	41.87	7.84%	6407
82 Tønsberg/Horten	Vestfold	41.86	11.93%	7742
83 Trondheim	Sør-Trøndelag	41.74	9.88%	17720
84 Holmestrand	Vestfold	41.59	10.37%	1173
85 Levanger/Verdalsøra	Nord-Trøndelag	41.56	8.59%	3457
86 Elverum	Hedmark	41.41	10.28%	3600
87 Rjukan	Telemark	41.27	6.00%	578
88 Hønefoss	Buskerud	40.90	9.27%	3117
89 Risør	Aust-Agder	39.47	7.83%	1048

Tabell A.2: Askeladd-indeks

	Kommune	Fylke	Andel askeladder
1.	Fjell	Hordaland	20.57%
2.	Stord	Hordaland	20.00%
3.	Asker	Akershus	19.01%
4.	Sola	Rogaland	17.32 %
5.	Karmøy	Rogaland	16.71%
6.	Stavanger	Rogaland	16.28%
7.	Lørenskog	Akershus	16.08%
8.	Oppegård	Akershus	15.62%
9.	Haugesund	Rogaland	15.31%
10.	Bærum	Akershus	15.27%
41.	Fredrikstad	Østfold	8.67%
42.	Ringerike	Buskerud	8.42%
43.	Nedre Eiker	Buskerud	8.31%
44.	Stange	Hedmark	8.16%
45.	Sarpsborg	Østfold	7.99%
46.	Gjøvik	Oppland	7.97%
47.	Ringsaker	Hedmark	7.75%
48.	Rana	Nordland	7.67%
49.	Skien	Telemark	6.65%
50.	Stjørdal	Nord-Trøndelag	6.39%

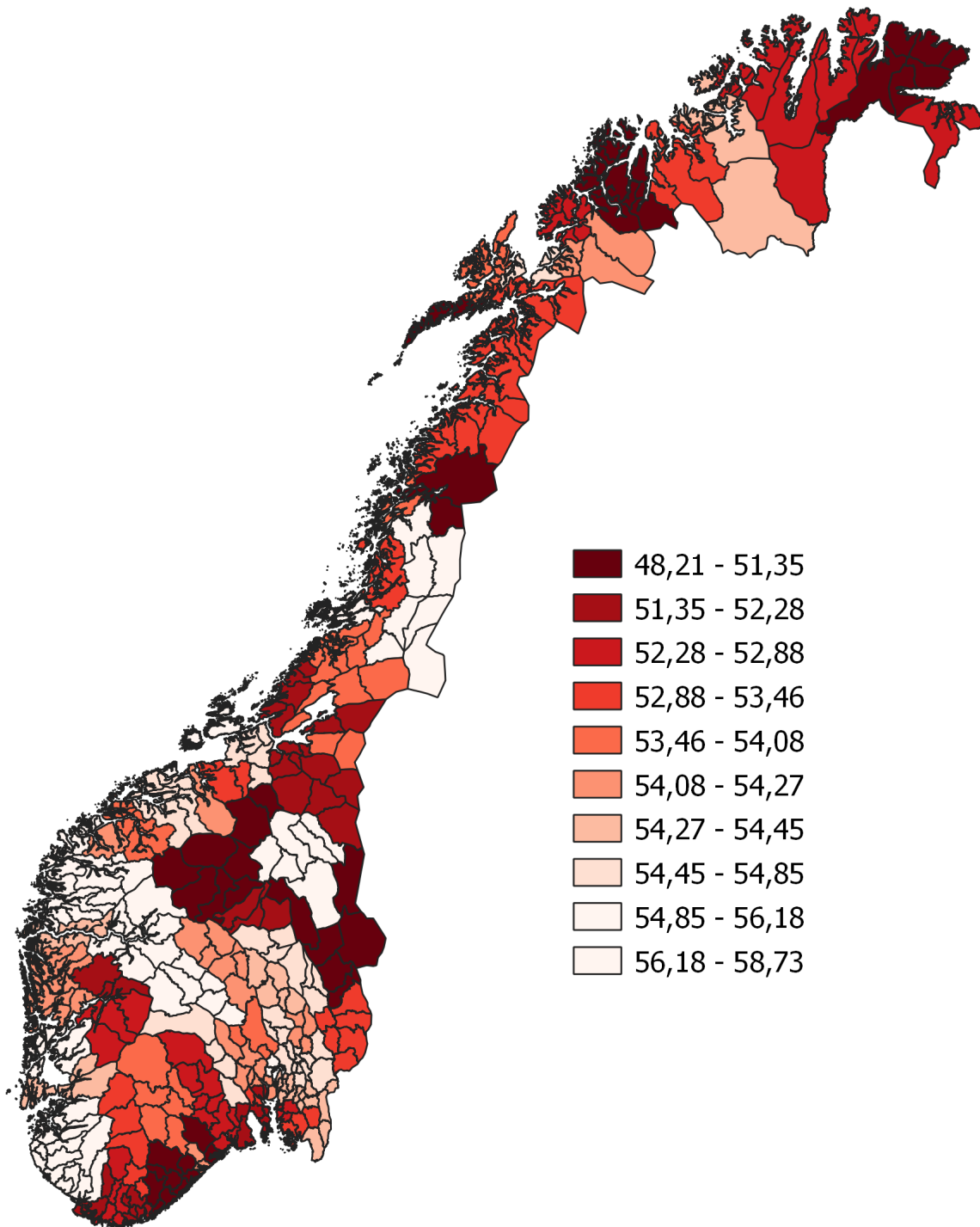
Noter: Andelen av barna som vokser opp hos far med inntekt i den laveste kvintilen av den nasjonale inntektsfordelingen som selv havner i den høyeste kvintilen. Avgrenset til å se på de 50 kommunene med flest observasjoner i utvalget, og viser de 10 kommunene med høyest andel og 10 kommunene med lavest andel.

Figur A.1: Estimert på b_m med inntekt ved ulike aldre



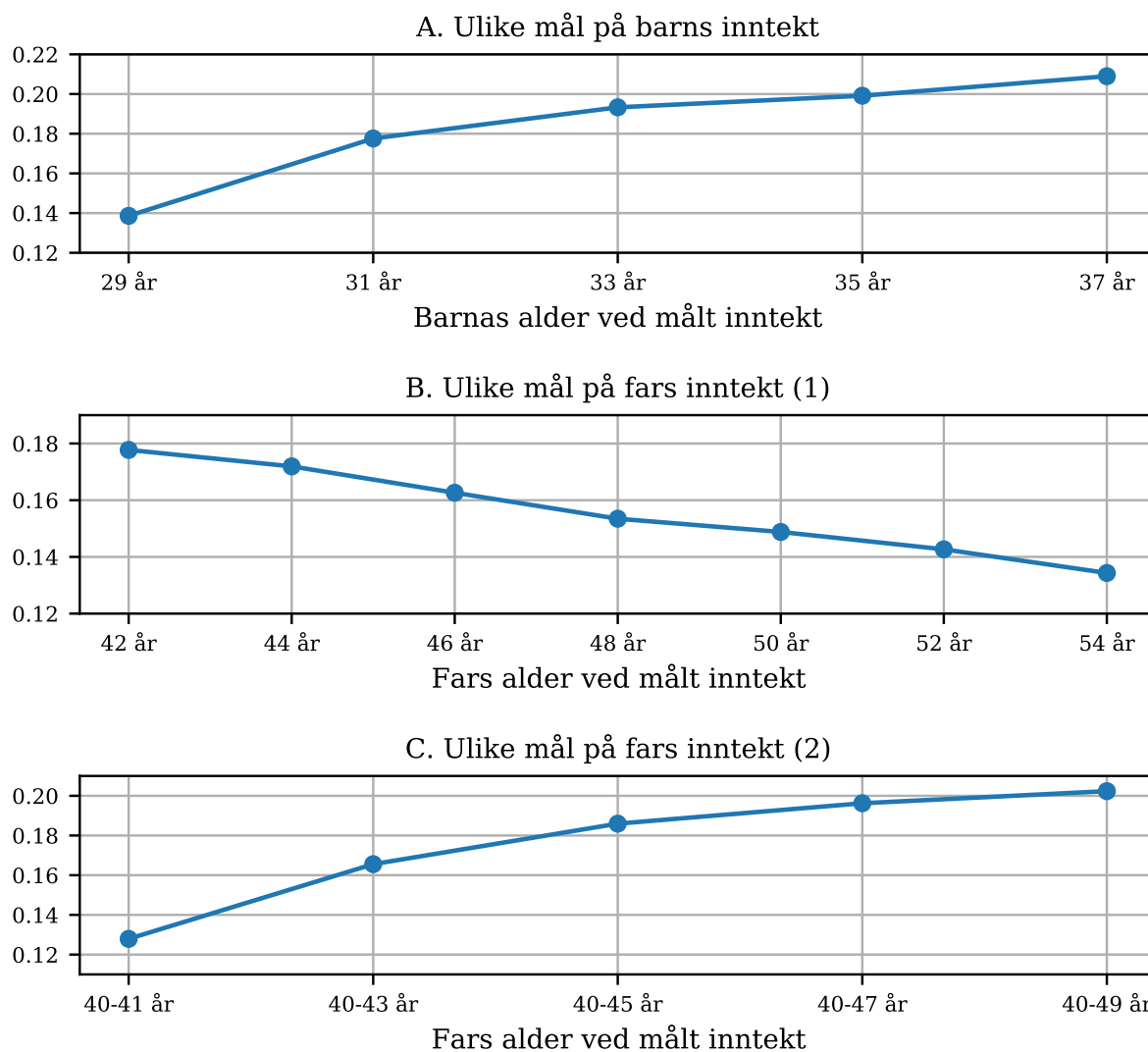
Noter: Figuren viser de estimerte koeffisientene av $\{b_m\}$ fra spesifikasjon (8.1) samt helning fra regresjon av b_m på flyttealder m , vektet med antallet observerte flyttere i de ulike aldre. Utvalget er avgrenset av barna som er født mellom 1977-1984 og består av barna som flytter én gang mellom 150 største kommunene i alderen 8-17 år. I motsetning til figur 8.1 er det samme avgrensning av kohortene med inntekt målt i de ulike aldre.

Figur A.2: Absolutt mobilitet i øverste halvdel av inntektsfordelingen



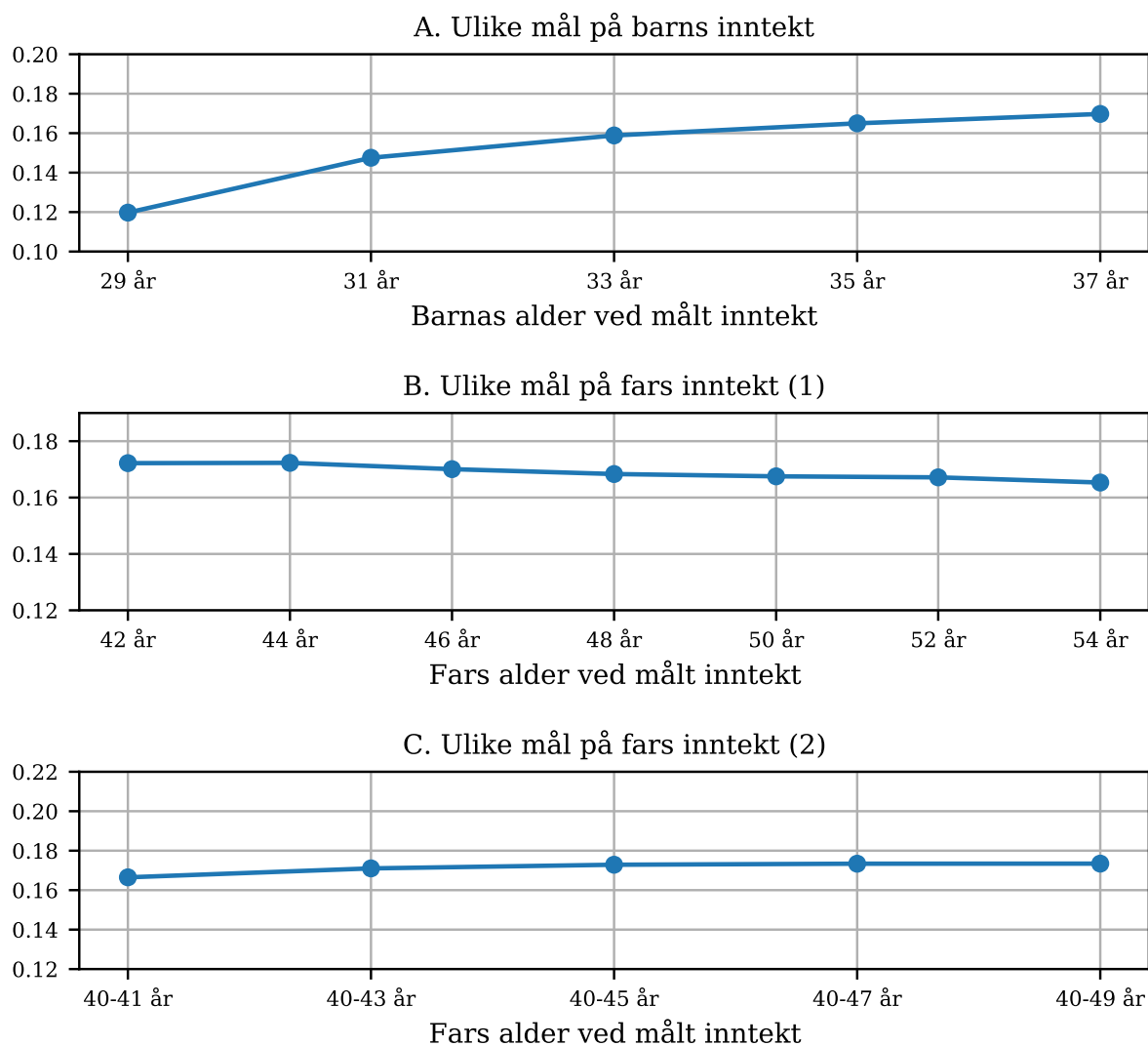
Noter: Kartet viser estimatene på \bar{y}_{75} fra ligning (7.5) i de ulike økonomiske regionene. Estimaten er delt i 10 like store grupper ut fra antall kommuner. Høyere estimat vises med lysere farge og medfører høyere absolutt mobilitet for barna som vokser opp med far i øverste halvdel av inntektsfordelingen. Estimaten er basert på kjerneutvalget og barna er plassert i den økonomiske regionen de hadde folkeregistrert adresse da de var 16 år.

Figur A.3: IGE-estimat med ulike mål på barns inntekt



Noter: I panel A er inntekt til far målt som gjennomsnitt av inntekt i alderen 40-49 år og inntekt til barna er målt som gjennomsnitt av et intervall på tre år sentrert i de angitte aldre. I både panel B og C er inntekt til barna målt som gjennomsnitt av inntekten i alderen 34-38 år. Panel B måler inntekt i et intervall på 5 år sentrert i de angitte aldre og panel C måler inntekt til far i de angitte aldre. Estimaten av IGE er stigende i C fordi inntekt målt i lengre intervall medfører mindre klassisk målefeil og dermed mindre forventningsskjevhet nedover. Inntekt er avgrenset til pensjonsgivende inntekt.

Figur A.4: Korrelasjon av inntektsrangering med ulike mål på inntekt



Note: Bruker tilsvarende mål på inntekt som i figur A.3. I motsetning til i hovedanalysen er kvantilene beregnet etter at far blir koblet på barn. Lengdene av intervallene på y-aksen tilsvarer figur A.3 for å gjøre det enklere å sammenligne forskjeller i sensitiveten til målene på hvilke aldre inntekt til far og barn blir målt.

A.1 Klassisk målefeil

Vi vil estimere

$$y_i = \alpha + \beta x_i + \epsilon_i \quad (\text{A.1})$$

der y logaritmen av *permanent inntekt* til barnet, x er tilsvarende for foreldrene og ϵ er feilledd der $\text{cov}(x, \epsilon) = 0$. Hvis vi hadde observert (y, x) kunne vi konsistent estimert β med OLS, men i praksis har vi ofte kun observasjon av inntekt i et begrenset tidsintervall. Dette medfører to kilder til målefeil. Den første er klassisk målefeil som følge av tilfeldige avvik fra gjennomsnittlig livsinntekt i intervallet vi observerer. Det andre er livssyklusskjevhet som følge av at inntekt gjennom livsløpet følger en konkav løpebane. Dette problemet er forsterket av at inntekten ulike yrkesgrupper har ulike løpebaner, der inntekten i yrker som krever mindre utdanning gjerne stagnerer tidligere. Dersom inntekt til barna blir målt for tidlig vil det derfor kunne overvurdere grad av mobilitet.

Vi ser kun på klassisk målefeil og antar derfor at den observerte inntekten til foreldrene som vi observerer på et gitt tidspunkt kan uttrykkes som

$$\tilde{x}_i = x_i + u_i \quad (\text{A.2})$$

der u_i er tilfeldig støy som er ukorrelet med de andre variablene i modellen, $\text{cov}(u_i, x_i) = \text{cov}(u_i, y_i) = \text{cov}(u_i, \epsilon_i) = 0$. Vi kan se bort i fra målefeil av inntekt til barna siden det kun påvirker presisjon til estimatoren. Substituerer (A.2) inn i (A.1):

$$\begin{aligned} y_i &= \alpha + \beta(\tilde{x}_i - u_i) + \epsilon_i \\ &= \alpha + \beta\tilde{x}_i + (\epsilon_i - \beta u_i) \end{aligned} \quad (\text{A.3})$$

der det nye feilleddet er $\epsilon_i - \beta u_i$. Merk at $\text{cov}(\tilde{x}_i, u_i) > 0$ fra (A.2). Det medfører at OLS på (y, \tilde{x}) vil være en inkonsistent estimator for β . OLS-estimatoren konvergerer i sannsynlighet til

$$\begin{aligned} \text{plim}(\hat{\beta}) &= \frac{\text{cov}(\tilde{x}_i, y_i)}{\text{var}(\tilde{x}_i)} = \frac{\text{cov}(x_i + u_i, \alpha + \beta x_i + \epsilon_i)}{\text{var}(x_i + u_i)} \\ &= \frac{\text{cov}(x_i, \alpha + \beta x_i + \epsilon_i)}{\text{var}(x_i + u_i)} + \frac{\text{cov}(u_i, \alpha + \beta x_i + \epsilon_i)}{\text{var}(x_i + u_i)} \\ &= \frac{\text{cov}(x_i, \alpha + \beta x_i + \epsilon_i)}{\text{var}(x_i + u_i)} = \beta \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} < \beta \end{aligned} \quad (\text{A.4})$$

Klassisk målefeil medfører dermed at IGE blir undervurdert (og tilsvarende at grad av sosial mobilitet blir overvurdert). Problemet kan dempes ved å se på gjennomsnittsinntekt

over et tidsintervall. For hver tidspunkt observerer vi

$$\tilde{x}_{it} = x_i + u_{it} \quad (\text{A.5})$$

Gjennomsnittlig observert inntekt er da:

$$\bar{\tilde{x}}_i = \frac{1}{T} \sum x_i + \frac{1}{T} \sum u_{it} = x_i + \bar{u}_i \quad (\text{A.6})$$

Anta nå at avvikene på ulike tidspunkt er ukorrelerte slik at

$$\text{var}(\bar{u}_i) = \text{var}\left(\frac{1}{T} \sum u_{it}\right) = \frac{\sigma_u^2}{T} \quad (\text{A.7})$$

Da vil

$$\begin{aligned} \text{plim}(\hat{\beta}) &= \frac{\text{cov}(\bar{\tilde{x}}_i, y_i)}{\text{var}(\bar{\tilde{x}}_i)} \\ &= \beta \frac{\sigma_x^2}{\sigma_x^2 + \frac{\sigma_u^2}{T}} \rightarrow \beta \end{aligned} \quad (\text{A.8})$$

når $T \rightarrow \infty$. Det er nok lite realistisk at avvikene kun er hvit støy. Vi kan åpne for persistens i avvik over tid ved å modellere dem som en første ordens autoregressiv prosess, AR(1), der

$$u_{it} = \rho u_{it-1} + \xi_t \quad (\text{A.9})$$

der $|\rho| < 1$ og $(\xi)_t$ er hvit støy. Det kan da vises at OLS på $(\bar{\tilde{x}}_i, y_i)$ konvergerer i sannsynlighet til

$$\text{plim}(\hat{\beta}) = \lambda_T \beta \quad (\text{A.10})$$

der

$$\lambda_T = \frac{\sigma_x^2}{\sigma_x^2 + \gamma \frac{\sigma_u^2}{T}} \quad (\text{A.11})$$

og der

$$\gamma = 1 + 2\rho \frac{T - \frac{1-\rho^T}{1-\rho}}{T(1-\rho)} \quad (\text{A.12})$$

Merk at skaleringsfaktoren λ_T avhenger av T og ρ . Vi vil at $\gamma \frac{\sigma_u^2}{T} \rightarrow 0$ slik at $\lambda_T \rightarrow 1$ og $\text{plim}(\hat{\beta}) \rightarrow \beta$. Persistens til avvikene ($\rho > 0$) fører til at $\gamma > 1$ og det tar lenger tid før et gjennomsnitt av målinger utjevner målefeilene.

A.2 Utforming av script i microdata.no

Databehandlingen og den statistiske analysen er gjennomført i plattformen microdata.no som benytter en Stata-liknende syntax. Fremgangsmåten min for å lage scriptene i microdata har vært å generere de gjennom egne script i Python. Det er flere fordeler med dette. For det første er det mulig å utforme scriptet slik at utvalgsavgrensning (og lignende) kan endres ved å kun endre parametre på begynnelsen av scriptet. For det andre kan de også redusere antallet linjer med koder som brukeren må skrive siden microdata ikke har løkkefunksjonalitet. For det tredje blir det enklere å bruke tall fra tabeller utenfor microdata i scriptet, for eksempel en tabell over konsumprisindeksen for å justere monetære størrelser for inflasjon. Ved å automatisere denne prosessen reduseres sannsynligheten for brukerfeil.

I hovedanalysen har jeg i hovedsak brukt to script: et deskriptivt som script som dokumenter geografiske forskjeller i inntektsmobilitet og et script for flyttedelen. Jeg skal nå beskrive utformingen av disse i mer detalj.¹

A.2.1 Deskriptiv

Jeg begynner med å avgrense utvalget av barna til individer født i de angitte årene og som er bosatt i Norge da de var 16 år samt i årene jeg måler deres inntekt. Deretter importerer jeg inntektsmål for ulike år og justerer de for inflasjon. Differansen mellom året inntekt er målt og året barna er født vil da angi alderen til barna på det tidspunktet. Dette kan brukes til å lage målet på inntekt. Deretter lager jeg et nytt datasett og gjør det samme for å avgrense utvalget av fedre og beregne deres inntekt. Siden barna har en variabel med fødselsnummer til far vil det deretter være mulig å koble variabler fra fedrene på deres respektive barn.

Det er nå enkelt å beregne nasjonale estimat. De regionale estimatene kan være mer utforderende siden det er mange geografiske enheter. Et smart triks her er å bruke interkassjonsledd for å få ut alle estimatene i én regresjonstabell i stedet for å kjøre regresjonene separat for hvert sted. Eksempel der variabelen bosted angir for eksempel kommune eller økonomisk region:

```
regress prosentil_barn c.prosentil_far##i.bosted , noconstant
```

Dette vil gi konstant for hvert sted, en helning og differansen fra denne helningen for de ulike stedene. Ved å behandle denne tabellen i annen programvare vil det da være mulig å finne konstant og helning for hvert sted. Det er også mulig å finne andelen som går fra nederste til øverste kvintil ved å definere en indikatorvariabel for om barna er i øverste kvintil og kjøre en regresjon:

```
generate kvintil_far = quantile(inntekt_far,5)
```

¹Jeg forsøker å forklare logikken i scriptene og bruker derfor til dels pseudokode. Se selve scriptene for korrekt syntax: <https://github.com/SverreFL/masteroppgave>

```

generate kvintil_barn = quantile(inntekt_barn,5)
generate topp_kvintil = kvintil_barn==4
regress topp_kvintil i.bosted if kvintil_far == 0, noconstant

```

En annen utfordring er at det ikke er en variabel for økonomiske regioner slik at denne må omkodes fra kommunenummer.

A.2.2 Flyttedelen

Dette scriptet er litt mer sammensatt. Avgrensingen av utvalgene og genereringen av inntektsvariabelene skjer i utgangspunktet på lignende måte som over, men forskjellen er jeg nå har informasjon om bosted for hvert år fra 1992-2006 i stedet for kun bosted da barna var 16 år. Etter å ha koblet inntektskvantilene til fedrene på datasettet til barna må jeg finne ut hvem av barna som flyttet, hvor mange ganger de har flyttet, alderen de flyttet og hvor de flyttet fra og til. Jeg lager variabler som angir antallet ganger barna har flyttet, året de flytter og hvor de flytter til. Jeg avgrenser til 150 største kommunene og kjører jeg en løkke på formen:

```

keep if (kommune92 == 101) | ... | (kommune92 == 2002)
for t in [1993, ..., 2006]:
    replace flytter = flytter+1 if (kommune{t} != kommune{t-1}) & ({t} -
        kohort <= 18)
    replace flyttetidspunkt = {t-1} if (kommune{t} != kommune{t-1}) & (
        flytter == 1)
    replace destinasjon = kommune{t} if (flytter == 1) & ({t} - kohort ==
        18)
keep if (destinasjon == 101) | ... | (destinasjon == 2002)
generate flyttealder = flyttetidspunkt-kohort if flytter == 1

```

Scriptet blir forenklet av at jeg kun ser på delmengden av flyttere som kun flytter én gang. Destinasjonen blir da kommunen barnet er bosatt i da det var 18 år, som er siste året jeg registrer om de flytter. Opprinnelsen blir kommunen barna er bosatt i året 1992. I første omgang avgrenser jeg til de permanente innbyggerne som ikke flytter i årene jeg observerer før de blir 18 år, og estimerer rangmobiliteten på de ulike stedene:

```

keep if flytter == 0
regress prosentil_barn c.prosentil_far##i.kommune92, noconstant

```

Deretter kjører jeg scriptet på nytt fra begynnelsen, men denne gangen beholder jeg kun barna som flytter én gang. Jeg bruker de estimerte parametrene (a_c, b_c) for rangmobilitet på hvert sted c til å få finne gjennomsnittlig utfall til de permanentene innbyggerne med far i samme prosentil på opprinnelse og destinasjon.

```

keep if flytter == 1
for c in [101, ..., 2002]:
    replace y_po = a_{c}+b_{c}*prosentil_far if kommune92 == c
    replace y_pd = a_{c}+b_{c}*prosentil_far if destinasjon == c

```

```
generate delta = y_pd -y_po
```

Jeg har nå størrelsene jeg trenger for å estimere $\{b_m\}$ og γ . Denne siste utfordringer er å bruke familie fasteffekt i microdata, siden det ikke er mulig å lage panel med variabler som ikke har tidsforløp. Jeg lager i stedet variabler med avvik fra gjennomsnitt innad i familie. Det er mulig å finne gjennomsnitt i familien med kommandoen collapse. For å bruke dette til å finne avvik må jeg først klonе datasettet for også å beholde den opprinnelige variabelen.

```
clone-dataset script script1
use script1
collapse (mean) variabel , by (far_fnr)
rename variabel variabel_snitt
merge variabel_snitt into script on far_fnr
use script
generate variabel_avvik = variabel-variabel_snitt
```

Deretter kan jeg kjøre vanlig regresjon med variabler med avvik fra gjennomsnitt innad i hver familie. De rapporterte standardfeilene er vil være for lave fordi gjennomsnittene blir estimert utenfor prosedyren slik at den ikke benytter riktig antall frihetsgrader i estimeringen av variansen til feilledet. Det vil likevel være mulig å justere dette i etterkant. I utledningen nedenfor ser jeg bort fra at vi bruker opp en frihetsgrad på estimering av helningsparameteren og gjør den forenklede antagelsen at det i gjennomsnitt er to søsken per familie², slik

$$\hat{\sigma}_{MKM}^2 = \frac{1}{N}SSR \quad (\text{A.13})$$

$$\hat{\sigma}_{FE}^2 = \frac{1}{N(1-0.5)}SSR, \quad (\text{A.14})$$

der $\hat{\sigma}_{MKM}^2$ er estimatet fra minste kvadrats metode på avvik innad i hver familie og $\hat{\sigma}_{FE}^2$ er det riktige estimatet. Målet er å finne en skaleringsfaktor k slik at $\hat{\sigma}_{MKM}^2 \cdot k = \hat{\sigma}_{FE}^2$. Vi kan løse ligningen

$$\frac{1}{N}SSR \cdot k = \frac{1}{N(1-0.5)}SSR \quad (\text{A.15})$$

$$\implies k = \frac{1}{1-0.5} = 2, \quad (\text{A.16})$$

slik at $se(\widehat{\hat{\beta}}_{MKM}) \cdot \sqrt{2} = se(\widehat{\hat{\beta}}_{FE})$.

²Jeg gjør denne antagelsen i fordi korreksjonen ble gjort etter jeg mistet tilgang på microdata.no, slik at det ikke var mulig å finne eksakt antall familier