



Aalborg Universitet

Kandidatafhandling - Samfundsøkonomi

12. August 2019

Intergenerationel indkomstmobilitet

Metodiske udfordringer ved observationsbegrænsede longitudinale surveys

Anders Kallehauge
Kristian Lindhardt Nielsen

Vejleder:
Jesper Eriksen



AALBORG UNIVERSITET

Titelblad

Aalborg Universitet
Cand.oecon.
4. Semester

Institut for Økonomi og Ledelse

Fibigerstræde 2
DK-9220 Aalborg
www.business.aau.dk

Titel:

Intergenerationel indkomstmobilitet
-Metodiske udfordringer ved
observationsbegrænsede longitudinale surveys

Projektperiode:

Forårssemester 2019

Forfattere:

Anders Kallehauge (studienr. 20144861)
Kristian Lindhardt Nielsen (studienr. 20146867)

Vejleder:

Ph.D. Fellow - Jesper Eriksen

Antal normalsider: 66

Anslag inkl. mellemrum: 159.555

Afleveringsdato: 12. august 2019

Abstract

This master's thesis explores the methodological issues of examining intergenerational mobility in surveys limited by few observations. In most Nordic countries it is possible to examine intergenerational mobility with register data, however this is not the case for many other countries including USA and UK. The relevance of this master's thesis lies in explaining the methodological issues, so researches and students in the future know how to examine intergenerational mobility in longitudinal surveys. We first show the issues of using the most used measures in the literature of intergenerational mobility: intergenerational elasticity (IGE), rank-rank correlation and transition matrices with the Danish longitudinal survey DLSY (Danish Longitudinal Survey of Youth). We find that issues regarding few years of available data for income (Persistent Transitory Shocks) and measuring the income data at the right time (life cycle bias), are key contributors to bias when examining the intergenerational mobility in longitudinal surveys. Moreover, we show the limitations when using longitudinal surveys to examine the causal effects of intergenerational mobility. We do this with a case where we use a Difference-in-Difference (DiD) model and the Danish reform "Bistandsloven 1976" in DLSY. We find that the limitations of DLSY breaks fundamental assumptions of the DiD model, which makes an analysis of causal effects difficult. Using DLSY and Bistandsloven 1976 we are unable to create distinct control and treatment groups and examine the parallel trend assumption.

Our master's thesis contributes to the existing literature by combining information of methodological issues of examining intergenerational mobility with IGE, rank-rank correlation and transition matrices. Moreover, we are the first in the literature, to our knowledge, that outlines the issues of examining the causal effects using DiD in longitudinal surveys with limited observations.

Indhold

1. Indledning.....	1
2. Intergenerationel indkomstmobilitet	4
2.1. Teori	4
2.2. Metode	11
2.2.1. Intergenerationel elasticitet	11
2.2.2. Rank-Rank korrelation.....	14
2.2.3. Transitionsmatrice	15
2.3. Undersøgelser med surveys fra litteraturen	15
2.4. Data.....	20
2.5. Analyse	23
2.5.1. Intergenerationel elasticitet	24
2.5.2. Rank-rank korrelation	30
2.5.3. Transitionsmatrice	34
2.6. Sammenligning af DLSY og andre suverys	37
3. Kausale sammenhænge	38
3.1. Hvad siger litteraturen?.....	40
3.2. Tidlig uddannelse og intergenerationel mobilitet	41
3.3. Bistandsloven 1976 – En reform for tidlig uddannelse	43
3.4. Metode	44
3.4.1. Difference-in-differences	45
3.5. Datakrav for DiD i forhold til DLSY	49
4. Diskussion	54
5. Konklusion.....	61
Litteraturliste	64
Appendiks.....	69
Appendiks A	69

1. Indledning

Forskere, politikere og samfundsdebattører har i mange årtier været interesseret i ulighed. Især indenfor de sidste 5-10 år er emnet blevet populært, hvilket i høj grad skyldes ny forskning og udgivelser, som indikerer negative effekter ved ulighed (Piketty 2014; Dabla-Norris et al. 2015; Cingano 2014). Denne debat er langt fra afgjort og kan være svær at afgøre, eftersom det ofte er en diskussion præget af ideologi. Noget som de fleste dog kan blive enige om, er at det er godt at have et samfund, hvor der ikke er ulighed i muligheder. Lighed i muligheder kan forstås ved at familiebaggrund ikke har nogen betydning for opnået succes, givet hvor hårdt personen arbejder (Rawls, 1971). Af denne grund er det interessant for forskere at undersøge social mobilitet og herunder den intergenerationelle indkomstmobilitet. Adskillige forfattere har beskrevet hovedtendenserne i den økonomiske litteratur om intergenerationel mobilitet. Nyere eksempler er Black & Devereux (2011) og Torche (2015).

Den intergenerationelle indkomstmobilitet i et land er interessant i sig selv, men ved at sammenligne mobiliteten med andre lande opnår man en bedre forståelse for den observerede mobilitet. I praksis findes der flere forskellige metoder til at måle den intergenerationelle indkomstmobilitet, som alle har sine fordele og ulemper. I litteraturen bliver mobiliteten ofte målt ved den intergenerationelle elasticitet (IGE), rank-rank korrelationer eller transitionsmatricer. Overordnet set viser undersøgelserne i litteraturen at mobiliteten er højest i de nordiske lande og lavest i Storbritannien og USA (se f.eks. Jäntti et al., 2006). I undersøgelser af den intergenerationelle mobilitet opstår der dog ofte det problem, at resultaterne kan være påvirket af bias som følge af det tilgængelige data. I de nordiske lande bliver mobiliteten ofte undersøgt med administrativt data fra registre, fordi disse er tilgængelige i landene. Det samme er ikke tilfældet i mange andre lande som f.eks. Storbritannien, hvor det typisk undersøges med longitudinale surveys og tværsnitsdata med en begrænset mængde observationer. Brug af disse former for surveys har vist sig at have forskellige implikationer for det empiriske resultat af mobiliteten. Litteraturen viser bl.a., at få års observationer af indkomstdata og tidspunktet hvor indkomsten måles på, kan være med til at overestimere mobiliteten i samfundet (Mazumder 2005; Haider & Solon 2006).

En anden udfordring i litteraturen om intergenerationel mobilitet har været at påvise hvilke faktorer der kausalt påvirker mobiliteten (Black & Devereux, 2011). En kausal sammenhæng forklarer årsagen til, at en bestemt variabel bevæger sig. Hvis man kender de kausale sammenhænge, er det f.eks. muligt at indføre politiske tiltag, som påvirker variabelen til fordel for ens interesser. Det er derfor interessant at undersøge de kausale sammenhænge mellem intergenerationel mobilitet og relevante variable. I

litteraturen var der i starten stor fokus på, hvor meget genetik havde af betydning i forhold til den intergenerationelle mobilitet, mens der i nyere tid især er sat fokus på, hvordan forældre og offentlige politikere kan påvirke børnenes udfald. Dette kræver i høj grad smarte metodiske valg, som gør at de kausale effekter kan isoleres og kan i praksis være svært at udføre. Et eksempel på hvordan det kan undersøges, er udgivelser fra litteraturen, som undersøger den intergenerationelle mobilitet på adopterede børn (Sacerdote 2007; Plug 2004). Idéen med dette er at finde en gruppe individer, som har det samme udgangspunkt, men er blevet introduceret til forskellige miljøer. Det kræver dog, at det er tilfældigt hvilke familier de adopterede børn bliver tildelt og at de bliver behandlet på samme måde som biologiske børn. Dette er sjældent tilfældet i virkeligheden og kan derfor være svært at undersøge. En variabel som ofte bliver undersøgt i forbindelse med den intergenerationelle mobilitet er uddannelse. Det er i høj grad pga. den tætte sammenhæng mellem uddannelse og indkomst. I praksis kan sammenhængen mellem intergenerationel mobilitet undersøges på forskellige måder. Det kan f.eks. undersøges, hvilken betydning forældrenes uddannelsesniveau har på børnenes fremtidige indkomst. Et andet alternativ er at undersøge, hvordan en stigning i mængden af uddannelse hos børnene vil påvirke den intergenerationelle mobilitet. Dette kan bl.a. gøres med en difference-in-differences (DiD) model. Ved brug af denne økonometriske metode er det f.eks. muligt at undersøge hvilken effekt en tidligere reform har haft på en kohorte. Et eksempel på brugen af denne fremgangsmåde ses hos Pekkarinen et al. (2009). Forfatterne bruger metoden til at undersøge, hvordan en skolereform i Finland påvirkede den intergenerationelle mobilitet. Deres undersøgelse blev udført ved hjælp af registerdata og reformen blev indført på forskellige tidspunkter i landet, hvilket gjorde det muligt at sammenligne med kontrolgrupper. Samme undersøgelse kan være problematisk uden registerdata, når der er begrænsede mængder af observationer til at forme kontrolgrupper.

Udfordringer med estimation af kausale sammenhænge kan ofte opstå som følge af begrænsede mængder data. Den mest almindelige data tilgængeligt til undersøgelserne af intergenerationel mobilitet er longitudinale surveys og tværsnitsdata. Der findes dog lande, som har store administrative databaser heriblandt Danmark og de andre nordiske lande. Men eftersom dette ikke er tilfældet for langt de fleste lande i verden, er det vigtigt at forstå de problematikker der opstår, når man undersøger den intergenerationelle indkomstmobilitet med data fra et survey, som ofte har relativt få observationer.

I dette projekt ønsker vi at belyse, hvilke problematikker der opstår i undersøgelser af den intergenerationelle mobilitet og begrænsninger der opstår i praksis, når der anvendes observationsbegrænsede datasæt, eksempelvis fra longitudinale surveys. Vores problemformulering lyder:

Hvilke metodiske problematikker opstår der i forbindelse med at undersøge den intergenerationelle indkomstmobilitet med data fra typiske observationsbegrænsede longitudinale surveys?

For at besvare den overordnede problemformulering har vi opstillet to underspørgsmål, som vi vil besvare i løbet af denne afhandling.

Underspørgsmål:

1. Hvilke metodiske problematikker opstår der i forbindelse med at estimere den intergenerationelle mobilitet deskriptivt i observationsbegrænsede surveys (IGE, rank-rank og transitionsmatricer)?
2. Hvilke metodiske problematikker opstår der i forbindelse med at finde de kausale sammenhænge mellem den intergenerationelle mobilitet og tidlig uddannelse?

Det første spørgsmål vil blive besvaret i afsnit 2, hvor vi anvender det danske survey DLSY. For at kunne lave denne undersøgelse giver vi først en introduktion til den grundlæggende teori om intergenerationel mobilitet. Dette er med til at give os en forståelse af, hvorfor forældrenes indkomst kan have en indflydelse på børnenes indkomst. Herefter redegør vi for de mest anvendte metoder til at undersøge den intergenerationelle mobilitet, som er IGE, rank-rank korrelation og transitionsmatricer. For at kunne besvare vores problemformulering og underspørgsmål lægger vi især vægt på problematikkerne ved at bruge disse metoder. Derefter viser vi de mest anvendte surveys i USA og Storbritannien. Med undersøgelser fra litteraturen fremhæver vi de metodiske udfordringer ved at bruge disse surveys. Dette gøres for at få en fornemmelse af, hvordan andre surveys ser ud og hvilke problemer der generelt findes ud fra litteraturen. Dernæst introducerer vi det danske survey DLSY og laver en analyse af den intergenerationelle mobilitet ved hjælp af IGE, rank-rank korrelation og transitionsmatricer. I afslutningen af afsnittet sammenligner vi det danske survey DLSY, med de andre surveys og undersøgelser som blev introduceret tidligere. Dette giver os en forståelse for de generelle metodiske problematikker, som opstår i forbindelse med at måle den intergenerationelle mobilitet i surveys.

I afsnit 3 forsøger vi at svare på det andet underspørgsmål. Først laves der en gennemgang af litteraturen for at få et overblik af, hvordan de kausale sammenhænge undersøges i litteraturen. Herefter beskriver vi hvorfor, der kan være en sammenhæng mellem tidlig uddannelse og den intergenerationelle mobilitet med udgangspunkt i Heckman-kurven og forskning indenfor området. Dette gøres for at underbygge valget om at inddrage en reform, som påvirker tidlig uddannelse, hvilket er et

traditionelt eksempel på estimation af kausale sammenhænge fra litteraturen. Dernæst introducerer vi den udvalgte reform, som er Bistandsloven 1976 og efterfølgende redegør vi for den metodiske opsætning af en DiD model. Formålet med dette er at forstå, hvad det kræver at opstille modellen og dermed være i stand til at påvise metodiske problematikker, når metoden bruges med data fra et survey. Til sidst i afsnittet vil datakravene for DiD modellen blive sammenholdt med det danske survey DLSY for at udpege svaghederne ved at bruge et survey, når man skal finde de kausale sammenhænge for den intergenerationelle mobilitet. Afsnit 4 indeholder en diskussion af hvad resultaterne fra projektet kan bruges til og hvorfor de er relevante. Derudover indeholder afsnittet også en forklaring af begrænsningerne i projektet samt forslag til fremtidig forskning. Afslutningsvis vil problemformuleringen blive besvaret i konklusionen i afsnit 5.

2. Intergenerationel indkomstmobilitet

Dette afsnit har til formål at forklare begrebet intergenerationel indkomstmobilitet og vise hvilke metodiske problemer der kan opstå i en analyse af mobiliteten med det danske survey DLSY. Først redegøres der for det teoretiske grundlag, som er baseret på teoretiske modeller af Becker & Tomes (1979) og Solon (2004). Herefter forklares de mest anvendte metoder til at måle den intergenerationelle indkomstmobilitet samt problemerne, som opstår i forbindelsen med brugen af disse metoder. Efter dette introducerer vi kort de mest anvendte surveys indenfor litteraturen samt redegør for undersøgelser, som benytter disse surveys til at undersøge den intergenerationelle mobilitet. Dernæst præsenteres og diskuteres vores analyse af den intergenerationelle indkomstmobilitet i det danske datasæt DLSY. Herefter sammenholder vi vores analyse med de generelle undersøgelser fra litteraturen. Dette gøres for at opnå en bred forståelse af de generelle metodiske problematikker ved at undersøge den intergenerationelle indkomstmobilitet i observationsbegrænsede surveys.

2.1. Teori

I dette afsnit præsenterer vi den mest anvendte teoretiske litteratur indenfor intergenerationel indkomstmobilitet. Formålet med dette er at beskrive, hvilke sammenhænge der er mellem forældre og børns fremtidige indkomst. Afsnittet tager udgangspunkt i en kort gennemgang af en model udviklet af Becker & Tomes (1986). Herefter vises den udvidede model af denne, som er lavet af Solon (2004). Disse modeller er fundamentale i litteraturen (Black & Devereux, 2011).

Det generelle fokus ved fordeling af indkomst og formue lå i mange år på de forskelle/uligheder, der eksisterer mellem individer og familier. I 1979 udgav Gary S. Becker & Nigel Tomes i artiklen 'An

Equilibrium Theory of the Distribution of Income and Intergenerational Mobility’ en af de første økonomiske modeller, der søger at forklare uligheden mellem generationer internt i familier. Modellen er siden blevet videreudviklet af Becker & Tomes (1986) og Solon (2004).

Modellens formål er at beskrive overførslen af indkomst, aktiver og forbrug fra forældre til efterkommere. I modellen afhænger forældres nytte ikke alene af deres eget forbrug, men også af deres børns fremtidige forbrug. Forældrene antages derfor at være intergenerationelt nyttemaksimerende i den forstand, at de bekymrer sig om deres børns ve og vel. For at uddybe de grundlæggende elementer i modellen viser vi først udvalgte ligninger, som er fælles for Becker & Tomes (1986) og Solon (2004), der beskriver sammenhængen mellem forældrenes indkomst og børnenes fremtidige indkomst. Becker & Tomes (1986) beskriver børnenes indkomst ved:

$$I_{t+1} = a + bI_t + \varepsilon_{t+1} \quad (2.1)$$

Ligning 2.1 beskriver hvordan børns indkomst I_{t+1} afhænger af forældrenes indkomst I_t samt konstanterne a og b . De stokastiske effekter ε_{t+1} er uafhængigt af forældrenes indkomst.

Indkomstuligheden mellem generationerne vil vokse over tid så længe $b \geq 1$, mens uligheden vil nærme sig et stabilt niveau (i absolutte værdier) hvis $b < 1$. Det ses derfor at b angiver hvorvidt et barn fra en rig familie, vil være bedre eller værre stillet end forældrene, når man ser på indkomst. Det samme gør sig gældende for børn i lavere bemidlede familier. Becker & Tomes (1986) definerer derfor intergenerationel mobilitet således:

’The degree of regression toward or away from the mean in the achievements of children compared to those of their parents is a measure of the degree of equality of opportunity in a society’

Den resterende del af modellen har til formål at beskrive, hvad der påvirker parametrene i ligning 2.1. En første approksimation i modellen, omhandler efterkommeres evner og egenskaber. Ved evner og egenskaber forstås de genetiske og kulturelle elementer som giver børn fordele/ulemper. Disse egenskaber nedarves i modellen gennem DNA og familiens kultur, og det antages derfor i modellen, at forældre ikke direkte kan investere i dette. I modellen skelnes der ikke mellem genetiske og kulturelle egenskaber. Egenskaber og evner udtrykkes i ligning 2.2:

$$E_t^i = \alpha_t + hE_{t-1}^i + v_t^i, \quad (2.2)$$

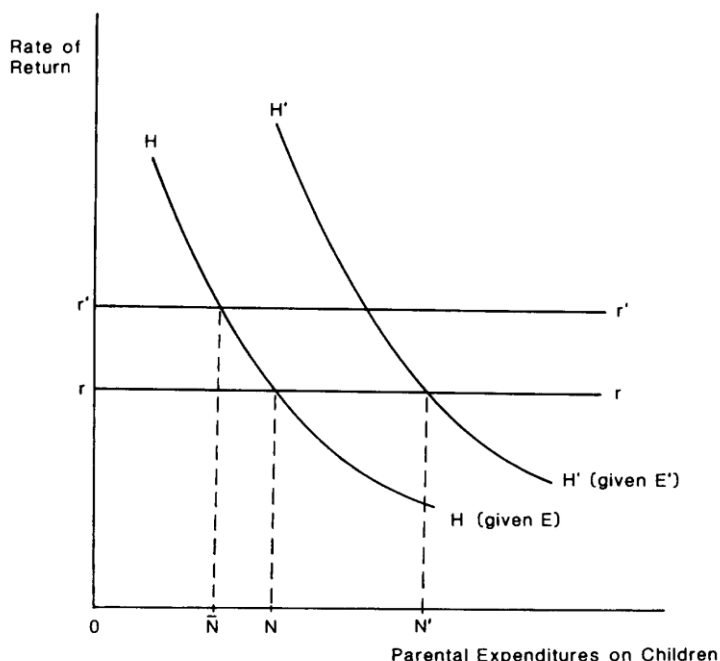
hvor E_t^i er egenskaber og evner ved familien i for generation t . Ved h gives arveraten, dvs. personers evne til at videreføre evner og egenskaber fra tidligere generationer. Det antages, at efterkommere kun delvist kan aftage deres forældres egenskaber dvs. $0 < h < 1$.

Antagelsen siger derfor, at evner og egenskaber bevæger sig mod gennemsnittet på tværs af generationer. Børn med højtbegavede forældre, vil også selv have evner og egenskaber der ligger over gennemsnittet, men stadig mindre end forældrene. Dette gælder ligeledes for børn med forældre, der er under gennemsnittet. a_t angiver de evner og egenskaber som er fælles for alle personer i et samfund. Denne er ikke konstant, eftersom staten kan vælge at investere i de socialfælles evner og egenskaber.

Det er ikke kun genetiske og kulturelle egenskaber som forældre har mulighed for at give videre. I modellen kan forældre påvirke deres børns fremtidige indkomst ved at investere i deres færdigheder, sundhed, indlæring og motivation samt en lang række andre egenskaber. Mængden af udgifter som forældre bruger på at fremme deres børns ve og vel, afhænger af forældrenes præferencer og indkomst, samt de offentlige udgifter til børns uddannelse og andet human kapital. Eftersom lønindtægter er den primære indtægt for de fleste personer, er det også den vej igennem, at forældre forsøger at påvirke deres børns succes i livet, dvs. deres fremtidige lønindkomst.

Den optimale mængde af investering er givet ved skæringen mellem den horisontale linje rr og den nedadgående kurve HH eller $H'H'$ i figur 2.1. N angiver mængden af investering i børnene, H angiver human kapital hos børnene (hvor H' er mængden af human kapital hos børn med højere egenskaber E') og rr angiver det marginale afkast i forhold til renten (det antages at forældrene kan låne til at investere i børnene). Figuren indeholder rate of return på y-aksen og forældrenes investering i børnene på x-aksen, fordi der i modellen antages, at forældrene vil maksimere børnenes velfærd (så længe det ikke sænker deres eget forbrug). Forældrene kan i modellen låne på perfekt fungerende kapitalmarkeder til at investere i børnene. De forskellige skæringspunkter i figuren viser ligevægtsinvesteringer fra forældrene på baggrund af at børnene kan betale lånene til investeringen tilbage. Yderligere ses det i figur 2.1, at børn med bedre egenskaber akkumulerer mere human kapital. I modellen antages det, at en relativ højere mængde human kapital vil medføre en højere forventet indkomst. Derudover ses det også, at mængden af investering i human kapital vil falde, hvis renten stiger, eftersom det vil påvirke forældrenes evne til at låne til investeringer i børnene.

Figur 2.1 – Optimal investering i børns human kapital



Kilde: Becker & Toms 1986 – Figuren viser den optimale mængde investering i børn, hvor H er human kapital og H' er human kapital for børn med højere egenskaber E' . N angiver mængden af investering og r angiver det marginale afkast i forhold til renten.

Solon (2004) modificerede Becker & Toms' model med henblik på at tilpasse den til nyere litteratur (Reville 1995; Fertig 2001), som fokuserer på den intergenerationelle elasticitet. Overordnet set viser modellen, at elasticiteten afhænger af fire faktorer: styrken af transmissionen af de "mekaniske" indkomstgenererende egenskaber (f.eks. genetik), effektiviteten af investering i børns human kapital, afkastet i indkomst til human kapital og progressiviteten af offentlig investering i børns human kapital. Dette betyder, at hvis land A har en mindre intergenerationel indkomstmobilitet end land B, kan det være på grund af en stærkere heritabilitet, mere produktive investeringer i human kapital, højere afkast til human kapital eller mindre progressive offentlige investeringer i human kapital (Solon, 2004). Vi beskriver her den grundlæggende opbygning af modellen.

Det antages at familie i indeholder en forælder af generation $t - 1$ og et barn af generation t . Familien skal allokere forælders livsindkomst efter skat, $(1 - \tau)y_{i,t-1}$, mellem forælders forbrug $C_{i,t-1}$ og investeringer i barnets human kapital, $I_{i,t-1}$. Dette giver budgetbegrænsning beskrevet i ligning 2.3.

$$(1 - \tau)y_{i,t-1} = C_{i,t-1} + I_{i,t-1} \quad (2.3)$$

Det antages at alle skatteindtægter går til investeringer, $G_{i,t-1}$ i børns human kapital, som defineres ved ligning 2.4.

$$h_{it} = \theta \log(I_{i,t-1} + G_{i,t-1}) + e_{it} \quad (2.4)$$

hvor, $G_{i,t-1}$ er regeringens investering i barnets human kapital (uddannelse eller sundhed), og θ er marginalproduktet af log-investeringen i human kapital, hvor logfunktionen indikerer et faldende marginalt produkt.

Mængden af human kapital et barn akkumulerer uafhængigt af forældre og regeringen er e_{it} . Becker & Tomes (1979) beskriver, at denne mængde er bestemt af *"the reputation and 'connections' of their families, the contribution to the ability, race, and other characteristics of children from the genetic constitutions of their families, and the learning, skills, goals, and other 'family commodities' acquired through belonging to a particular family culture. Obviously, endowments depend on many characteristics of parents, grandparents, and other family members and may also be culturally influenced by other families."*

Ud fra denne beskrivelse må det antages, at mængden af barnets e_{it} er positivt korreleret med forældrens mængde e_{it-1} . Derfor skrives funktionen som en første ordens autoregressiv proces som ligning 2.5.

$$e_{it} = \delta + \lambda e_{it-1} + v_{it} \quad (2.5)$$

Hvor, v_{it} er et white noise fejllid og λ er en heritabilitetskoefficient som ligger mellem 0 og 1.

Til sidst defineres barnets indkomst y_{it} ved ligning 2.6.

$$\log y_{it} = \mu + p h_{it} \quad (2.6)$$

Hvor p er afkastet i indkomst til human kapital.

Herefter udledes familiens investeringsadfærd. Først defineres forældrenes nyttefunktion ved en Cobb-Douglas nyttefunktion i ligning 2.7.

$$U_i = (1 - \alpha) \log C_{i,t-1} + \alpha \log y_{it} \quad (2.7)$$

α er en parameter for altruisme og ligger mellem 0 og 1. Den beskriver forældrenes præference mellem deres forbrug og børnenes indkomst.

Denne funktion kan omskrives ved at indsætte ligning 2.3 og 2.6 i ligning 2.7. Dette giver os ligning 2.8.

$$U_i = (1 - \alpha) \log[(1 - \tau)y_{i,t-1} - I_{i,t-1}] + \alpha\mu + \alpha\theta p \log(I_{i,t-1} + G_{i,t-1}) + \alpha p e_{it} \quad (2.8)$$

Den nye funktion indeholder nu kun en beslutning for forældrene, nemlig $I_{i,t-1}$. Den optimale investeringsadfærd kan beskrives ved førsteordensbetingelsen for forældrenes nyttemaksimeringsproblem.

$$\frac{\partial U_i}{\partial I_{i,t-1}} = \frac{-(1-\alpha)}{[(1-\tau)y_{i,t-1} - I_{i,t-1}]} + \frac{\alpha\theta p}{(I_{i,t-1} + G_{i,t-1})} = 0 \quad (2.9)$$

Ligning 2.9 løses for det optimale valg af $I_{i,t-1}$ og giver os ligning 2.10.

$$I_{i,t-1} = \left[\frac{\alpha\theta p}{1-\alpha(1-\theta p)} \right] (1 - \tau)y_{i,t-1} - \left[\frac{1-\alpha}{1-\alpha(1-\theta p)} \right] G_{i,t-1} \quad (2.10)$$

Ud fra ligning 2.10 kan flere ting udledes i forhold til investering i barnets human kapital: hvis offentlig investering, $G_{i,t-1}$, holdes konstant, vil forældre med højere indkomst investere mere i deres børns human kapital. Hvis skatter holdes konstant, vil højere offentlig investering i et barns human kapital til dels fortrænge forældrenes investering. Investeringen i barnets human kapital stiger med forældrenes altruisme (α). Til sidst kan det udledes, at forældrene har større incitament for at investere i barnets human kapital, når der er større udbytte. Dette kan ses ved, at forældrenes investering stiger i θp , som er afkastet i indkomst til human kapital.

Ovenstående resultater og antagelser bruges nu af Solon (2004) til at forklare den intergenerationelle sammenhæng mellem børnene og forældrenes indkomst. Ved at substituere ligningen for barnets human kapital (ligning 2.4) ind i ligningen for barnets indkomst (ligning 2.6) fås ligning 2.11.

$$\log y_{it} = \mu + p[\theta \log(I_{i,t-1} + G_{i,t-1}) + e_{it}] \quad (2.11)$$

Herefter indsættes ligning 2.10 i stedet for $I_{i,t-1}$ og ligningen omskrives til ligning 2.12.

$$\log y_{it} = \mu + \theta p \log \left[\frac{\alpha\theta p}{1-\alpha(1-\theta p)} \right] + \theta p \log[(1 - \tau)y_{i,t-1} - G_{i,t-1}] + p e_{it} \quad (2.12)$$

Skrevet på en anden måde giver dette ligning 2.13.

$$\log y_{it} = \mu + \theta p \log \left[\frac{\alpha\theta p(1-\tau)}{1-\alpha(1-\theta p)} \right] + \theta p \log \left\{ y_{i,t-1} \left[1 + \frac{G_{i,t-1}}{(1-\tau)y_{i,t-1}} \right] \right\} + p e_{it} \quad (2.13)$$

Denne omskrives med forudsætningen, at $\frac{G_{i,t-1}}{(1-\tau)y_{i,t-1}}$ er lille til ligning 2.14.

$$\log y_{it} = \mu + \theta p \log \left[\frac{\alpha \theta p (1-\tau)}{1-\alpha(1-\theta p)} \right] + \theta p \log y_{i,t-1} + \theta p \left\{ \left[\frac{G_{i,t-1}}{(1-\tau)y_{i,t-1}} \right] \right\} + p e_{it} \quad (2.14)$$

Ud fra denne ligning kan det ses, at den intergenerationelle mobilitet bliver påvirket af regeringens politik for offentlig investering i børns human kapital. Denne politik kan approksimeres ved:

$$\frac{G_{i,t-1}}{[(1-\tau)y_{i,t-1}]} \cong \varphi - \gamma \log y_{i,t-1} \quad (2.15)$$

Her vil en positiv værdi af γ signalere en relativ progressivitet i offentlig investering af børnenes human kapital. Den absolutte offentlige investering er ikke nødvendigvis højere for højindkomstfamilier, men ratioen af offentlig investering til forældreindkomst efter skat vil falde på samme tid med, at indkomsten falder for forældrene. Samtidigt kan det ses, at en højere positiv værdi for γ vil øge progressiviteten af politikken.

Til sidst bliver ligning 2.15 indsat i ligning 2.14, hvilket giver ligning 2.16.

$$\log y_{it} \cong \mu^* + [(1-\gamma)\theta p] \log y_{i,t-1} + p e_{it} \quad (2.16)$$

Ligning 2.16 minder om den samme ligning, som ofte bliver brugt i litteraturen til at måle den intergenerationelle indkomstmobilitet. I dette tilfælde vil fejleddet ikke kunne karakteriseres som iid, fordi det er korreleret med $\log y_{it}$ (e_{it} og $\log y_{i,t-1}$ er afhængig af $e_{i,t-1}$).

Herefter noterer Solon (2004) steady state for den intergenerationelle elasticitet, som er estimeret ofte brugt i den empiriske litteratur om intergenerationel indkomstmobilitet. Han henviser til Green (2000, s. 534-535), som viser, at steady state er summen af de to autoregressive parametre fra ligning 2.5 og 2.16, hældningskoefficienten fra ligning 2.16 og koefficienten af seriekorrelationen i ligning 2.5, divideret med 1 plus deres produkt giver ligning 2.17.

$$\beta = \frac{(1-\gamma)\theta p + \lambda}{1 + (1-\gamma)\theta p \lambda} \quad (2.17)$$

Ud fra dette kan vi undersøge sammenhængen mellem estimeret β fra ligning 2.17 og parametrene i modellen. Overordnet set vil den intergenerationelle elasticitet β stige sammen med parametrene λ , θ , p , og $1 - \gamma$. Den intergenerationelle elasticitet vil altså være større, hvis heritabiliteten er større (λ), hvis investeringerne i human kapital er mere produktive (θ), hvis afkastet i indkomst til human

kapital er større (p) eller hvis offentlige investeringer i børns human kapital er mindre progressive (γ er mindre).

Modellen angiver, at der er flere ting der kan påvirke den intergenerationelle indkomstmobilitet. Overordnet set viser modellen, at børnenes indkomst kan blive påvirket af deres evner til at arve egenskaber fra deres forældre og mængden af forældrenes investering i børnenes human kapital. Yderligere kan offentlige investeringer i børnenes human kapital påvirke deres indkomst. Ved at regressere børn og forældres logindkomst kan vi måle børnenes indkomst i forhold til forældrenes, men dette giver os ikke svaret på, hvilke ting der påvirker denne sammenhæng. Det er derfor umiddelbart svært at sige noget om, hvad det er der påvirker den intergenerationelle mobilitet ud fra β . Dette kræver mere komplekse metoder heriblandt en difference-in-differences model, som kan undersøge de kausale sammenhænge. Denne metode vil blive beskrevet i afsnit 3.4.

2.2. Metode

Dette afsnit har til formål at introducere de empiriske metoder, der anvendes i analysen af intergenerationel indkomstmobilitet samt problematikker og kritik af disse. Vi redegør i afsnittet for de mest anvendte metoder i litteraturen, som er intergenerationel elasticitet, rank-rank korrelation og transitionsmatricer. Vi vil især fokusere på problematikkerne, som er forbundet ved at bruge metoderne.

2.2.1. Intergenerationel elasticitet

En af de mest anvendte metoder til at måle intergenerationel indkomstmobilitet, er intergenerationel elasticitet (IGE). IGE findes ved en regression mellem børnenes indkomst og forældrenes indkomst i log. Dette giver os ligning 2.18.

$$\log(Y_1) = \alpha + \beta \log(Y_0) + \varepsilon \quad (2.18)$$

Hvor, $Y_1 = \text{indkomst børn}$, $Y_0 = \text{indkomst forældre}$, $\beta = IGE$

Parameteren β er altså målet for IGE. En højere koefficient indikerer en stærkere sammenhæng mellem børnenes indkomst og forældrenes. Mobilitetsmålet er invers i forhold til koefficienten, hvilket betyder, at en lavere koefficient er lig med højere mobilitet.

En af problematikkerne der opstår, når man skal måle IGE, er at definere den relevante indkomst for børnene og forældrene. Det antages typisk, at dette er den permanente indkomst. Den permanente indkomst defineres, som den forventede gennemsnitlige indkomst på lang sigt. Dette giver nogle udfordringer i forhold til at udregne individernes livsindkomst, eftersom mange datasæt ikke indeholder de nødvendige data for både børnene og forældrene.

I forlængelse af problematikken med permanent indkomst følger en udfordring med hvor mange år indkomstdata, der er til rådighed i datasættet. Tidlige estimater af mobiliteten anvendte indkomstdata fra ét år for både barn og forældre. Det blev dog påvist af arbejde udført af Solon (1992) og Zimmerman (1992), at der opstod en betydelig bias ved kun at bruge ét år. De forbedrede estimatet ved at bruge et gennemsnit af indkomsten fra 4-5 år. Nyere forskning har fokuseret på at opnå bedre estimater ved at bruge gennemsnittet fra endnu flere år for at tage højde for eventuelle choks til indkomsten og ved at fokusere på hvilket tidspunkt indkomstdata for barn og forældre er målt.

Vigtigheden af at bruge gennemsnittet for flere år beskrives i litteraturen ved "Persistent Transitory Shocks" (Black & Devereux, 2011). Problematikken kan beskrives matematisk. Vi antager følgende: $y_{0a} = y_0 + v_a$ hvor a er alderen på forældrene, når indkomsten er målt. Parameteren v er afvigelsen mellem den målte indkomst ved alderen a og den permanente indkomst, hvor det antages at v ikke er korreleret med y_0 . Hvis v udelukkende er et transitorisk shock, vil attenuation faktoren¹, når IGE måles med data fra ét år være som ligning 2.19. Attenuation faktoren er en skalering af regressionskoefficienten og beskriver dermed hvor meget koefficienten er nedskaleret. Den bruges f.eks. til at beskrive, hvor meget bias der er tilstede i mål af intergenerational mobilitet.

$$\frac{\text{var}(y_0)}{\text{var}(y_0) + \text{var}(v)} \quad (2.19)$$

Hvis man i stedet for tager gennemsnittet af indkomsten over T år, vil attenuation faktoren være ligesom ligning 2.20.

$$\frac{\text{var}(y_0)}{\text{var}(y_0) + \text{var}(v)/T} \quad (2.20)$$

Ud fra dette kan det ses, at attenuation faktoren bliver mindre jo flere år der inddrages i målingen af indkomst.

Mazumder (2005) påviser betydningen af den tilstedeværende persistens i de transitoriske shock ved hjælp af en AR(1) proces. Det antages at: $v_a = \delta v_{a-1} + \omega_a$ og ω_a er iid (independent and identically distributed). Dette giver os attenuation faktor som ligning 2.21.

$$\frac{\text{var}(y_0)}{\text{var}(y_0) + \gamma \text{var}(v)/T} \quad (2.21)$$

¹Hvis et estimat lider af attenuation bias betyder det, at parameteren vil blive underestimeret. Estimatet bliver altså lavere end den sande værdi af parameteren.

$$\text{Hvor, } \gamma = 1 + 2\delta \left(\frac{T - (1 - \delta^T)/(1 - \delta)}{T(1 - \delta)} \right)$$

Som det kan ses ud fra ligningerne, vil attenuation bias være større i forhold til eksemplet, hvor det udelukkende var et transitorisk shock hvis: $\delta > 0$. Dette er gældende for alle værdier af T . Denne pointe viser Mazumder (2005) ved hjælp af simulationer for forskellige værdier af δ og T samt benchmarkværdier for variansen af y_0 og v . Hvis $\delta = 0$ er det et udtryk for, at der ikke er seriekorrelation tilstede. Hans simulationer viser, at $\delta = 0.5$ giver en attenuation faktor på 0.69 med $T = 5$ år, hvilket indikerer, at estimatet underestimeres med omkring 30%. Derimod giver det en attenuation faktor på 0.83, når $\delta = 0$. Dette viser, at i tilfælde af tilstedeværelse af seriekorrelation, vil attenuation faktoren være lavere. På baggrund af en attenuation faktor på 0.69 foreslår han, at man burde opskalere estimerne af IGE i USA fra 0.4 til omkring 0.6 ($T = 5$). Desuden viser hans simulationer, at det kan kræve op til 20 eller 30 års indkomstdata for at få en attenuation faktor på 0.9 eller højere. Til sidst undersøger han på amerikansk administrativt data om IGE forandrer sig, når antallet af år stiger. Her finder han et estimat for IGE på 0.25 når $T = 2$, 0.45 når $T = 7$ og 0.61 når $T = 16$. Disse resultater samt førnævnte forklaringer indikerer altså, at IGE vil blive underestimeret, hvis der ikke bruges et gennemsnit af en lang årrække data for indkomst.

En anden problematik der kan opstå i målingen af IGE er ifølge litteraturen børnene og forældrenes alder, når deres indkomst måles. Mazumder (2005) beskriver denne problematik ved at undersøge variansen af det transitoriske element af indkomsten. Hvis variansen ændrer sig gennem livet og indkomsten er målt ved et punkt, hvor variansen er høj, vil det betyde, at estimer af IGE med få års data vil være biased. Baker & Solon (2003) og Mazumder (2005) viser, at variansen er lavest omkring 40-årsalderen. De viser, at kurven har en U-form, hvilket betyder, at attenuation bias i indkomstdata vil være lavere i 40-årsalderen i forhold til 30-årsalderen og 55-årsalderen. Det er derfor mest optimalt at måle indkomsten omkring 40-årsalderen.

En sidste problematik som er udbredt i litteraturen, er life cycle bias. Denne bias opstår, eftersom sammenhængen mellem nuværende indkomst og livstidsindkomst ikke er konstant over livscyklussen. Dette skyldes heterogene indkomstprofiler, alt efter hvilken alder et individ har. Individer med en høj livsindkomst vil som regel have en større vækst i indkomst end individer med en lavere livsindkomst. Forskellen mellem individerne med en lav livsindkomst og individerne med en høj livsindkomst, vil være lav tidligt i individernes indkomstprofil. Dette medfører, at forskellen i livsindkomsten vil blive underestimeret. Det er derfor vigtigt at overveje hvilken alder individerne har, når indkomsten måles

(Torche, 2015). I litteraturen findes der forskellige estimater af, hvornår indkomsten bør måles for at undgå life cycle bias. Haider og Solon (2006) estimerede sammenhængen mellem nuværende indkomst og permanent indkomst og påviste, at indkomsten burde måles tidligt i trediveårsalderen og til midten af fyrreårsalderen i USA. Landersø & Heckman (2017) underbyggede dette resultat for Danmark. Estimer fra Chetty et al. (2014) viser, at mobiliteten stabiliserer, når børn når slut-tyverne i USA, hvilket er betydeligt hurtigere end tidligere antaget.

Ud fra ovenstående metodiske problematikker er der noget der tyder på, at der højst sandsynligt vil opstå bias i målingen af IGE. Det er i høj grad muligt at tage højde for dette ved at inddrage gennemsnittet af en lang årrække af data for indkomsten og ved at måle indkomsten når aldersrelaterede bias er lavest. Eftersom det ikke altid er muligt at få adgang til det ”optimale” data, kan dette oftest ikke lade sig gøre.

2.2.2. Rank-Rank korrelation

En anden måde hvorpå den intergenerationelle indkomstmobilitet kan måles på, er ved hjælp af rank-rank korrelation. Denne metode bruges f.eks. af Chetty et al. (2014) til at måle den intergenerationelle indkomstmobilitet i USA. Rank-rank korrelationen findes ved at opdele børn og forældres indkomst i indkomstgrupper fra 0-100, hvor 0 er den lavest indkomstgruppe og 100 er den højeste. Herefter laves der en regression mellem børnene og forældrenes indkomst som kan ses i ligning 2.22:

$$R_i = \alpha + \beta_{PR}P_i + \varepsilon \quad (2.22)$$

Hvor, R_i = indkomstgruppe i for børnene, P_i = indkomstgruppe i for forældrene, β_{PR} = rank – rank korrelation

Rank-rank korrelationen måler sammenhængen mellem et barns position i indkomstgruppen i forhold til hans/hendes forældres position i indkomstgruppen. Målet er ligesom IGE invers i forhold til mobilitet. Dette betyder, at et større estimat er lig med lavere mobilitet og omvendt. I modsætningen til IGE, kan man med rank-rank korrelationer inddrage negative eller nulindkomster og kan derfor være et mere realistisk mål for mobiliteten. Derudover er målet også i nyere tid blevet en af de mere anvendte metoder, fordi det er tættere på linearitet i forhold til IGE. Et problem med forholdet mellem forældrenes og børnenes indkomst er, at de som regel ikke er lineære. Det viser sig dog, at rank-rank korrelationer som regel er tættere på linearitet og dermed er et mere præcist mål for mobiliteten (Chetty et al. 2014).

Eventuelle problematikker med bias er ikke så veludforskede, som for IGE. Et nyligt studie af Nybom og Stuhler (2016) undersøger eventuelle problematikker. Ligesom med IGE finder de en tendens til, at rank-rank korrelationen underestimeres. Biaset opstår på baggrund af samme målefejl som IGE. Derfor anbefales det at tage gennemsnittet af indkomsten for en lang årrække. Derudover påviser de også et life cycle bias, når rank-rank korrelationen måles. Det er dog i mindre grad end ved IGE. De foreslår derfor ligesom ved IGE at måle indkomsten, når individerne er tættest på deres permanente indkomst.

2.2.3. Transitionsmatrice

En tredje metode som også bruges i litteraturen til at måle den intergenerationelle indkomstmobilitet, er en transitionsmatrice. Denne metode kan ikke-parametrisk vise, hvor stor sandsynligheden er for et barn i en bestemt indkomstgruppe, flytter sig til en anden indkomstgruppe betinget af deres forældres indkomstgruppe. Børn og forældre kan f.eks. opdeles i fem kvintiler. I dette tilfælde kan vi undersøge sandsynligheden for, at et barn hvis forældre er fra den laveste indkomstkvin til, flytter sig til den højeste indkomstkvin til. I litteraturen bruges der ofte kvartiler eller kvintiler til at opdele indkomsten. Dette kritiseres af Bhattacharya og Mazumder (2011) eftersom, at det ifølge dem kan være svært at opdage mindre bevægelser i mobiliteten. De anbefaler derfor i stedet for at opdele indkomsten i percentiler.

Ligesom med IGE og rank-rank korrelation er der også problemer i forhold til at kunne måle individers livsindkomst, når der bruges transitionsmatricer. Denne problematik bliver forsøgt løst ligesom ved de andre metoder, hvor der bruges gennemsnittet af indkomsten over en lang årrække. O'Neill et al. (2007) undersøgte problematikken og viser, at klassiske målefejl i børnenes indkomst medfører en overestimering af mobiliteten, hvoraf den største bias finder sted i bunden af indkomstfordelingen (Black & Devereux, 2011)

2.3. Undersøgelser med surveys fra litteraturen

I dette afsnit vil vi redegøre for de mest anvendte surveys indenfor litteraturen af den intergenerationelle mobilitet. Den tilgængelige data varierer meget fra land til land. I litteraturen er den intergenerationelle mobilitet ofte undersøgt med registerdata i de nordiske lande, mens USA og Storbritannien ofte er undersøgt med surveys. Forskere har dog i nyere tid fået adgang til registerdata i USA. Vi vil i dette afsnit tage udgangspunkt i de mest anvendte surveys fra USA og Storbritannien samt præsentere eventuelle problemer, som forskere har haft i forbindelse med at bruge disse surveys til at

undersøge den intergenerationelle mobilitet. Formålet med dette er at få en bedre forståelse for, hvordan longitudinale surveys generelt ser ud og hvordan det danske surveydata fra DLSY repræsenterer det data. Derudover vil der også blive præsenteret en alternativ måde at bruge surveys, hvis de ikke indeholder informationer omkring forældrenes indkomst.

Den intergenerationelle mobilitet i USA er i mange år blevet undersøgt ved hjælp af surveys. De mest anvendte surveys er The Panel Study of Income Dynamics (PSID), The General Social Survey (GSS) og The National Longitudinal Surveys (NLS). PSID er et longitudinal panel survey af amerikanske familier, som er lavet af University of Michigans Survey Research Center. Undersøgelsen startede i 1968 og indeholder informationer omkring beskæftigelse, indkomst, formue, uddannelse osv. for 5000 familier i USA. Datasættet er eksempelvis blevet brugt af Mayer og Lopoo (2005) til at undersøge IGE. Et generelt problem som de fremhæver i deres undersøgelse, er usikkerhed i forhold til surveys. Denne problematik vil som udgangspunkt altid være til stede ved surveys, eftersom individerne ikke altid svarer 100 procent korrekt, når de bliver spurgt om indkomst. Derudover fremhæver de også, at deres estimer kan være påvirket af bias, eftersom de bruger et gennemsnit på syv år for forældrenes indkomst. Flere års gennemsnit end syv år ville medføre en sample size, der er for lille til at kunne bruges. Det anvendte indkomstdata for børnene, er et gennemsnit på tre år og er målt fra børnene bliver 30 år gamle. Deres baseline model indeholder 1537 observationer. Deres analyse giver statistisk signifikante resultater af IGE med relative små standardfejl. Estimerne kan dog stadig være påvirket af bias, eftersom de bruger relativt få års gennemsnit af indkomstdata. Udover at undersøge estimatet for IGE, undersøger de også hvordan trenden i IGE, har udviklet sig fra 1949 og frem til 1965. De undersøger trenden ved at opdele datasættet i 14 grupper således, at de har en gruppe fra 1949-1952, en fra 1950-1953 osv. Konklusionen af deres resultater er at selvom PSID som udgangspunkt, er den bedste kilde til at estimere trenden, kan de ikke drage stærke konklusioner fra undersøgelsen. Dette skyldes, at trenden ikke er monotonisk og deres sample size er lille.

NLS er fem surveys udført af bureauet for beskæftigelsesstatistik fra USA's beskæftigelsesministerie. Bloome & Western (2011) bruger NLS66 og NLSY79 til at undersøge IGE. NLS66 er et survey af 5.225 mænd i alderen 14 til 24, som interviewes i 1966 og herefter interviewes de regelmæssigt frem til 1981. NLSY79 er et survey af 12.868 mænd og kvinder i alderen 14 til 22 år i 1979. De interviewes første gang i 1979 og bliver herefter interviewet hvert år frem til 1994 og herefter hver halve år. I deres analyse undersøger de kun IGE for sønner, eftersom NLS66 kun indeholder informationer om mænd. De forsøger at undersøge eventuelle forskelle i race og opdeler derfor efter hvid og sort. Deres

model for IGE indeholder lidt over 1400 observationer for hvide og omkring 400 og 700 observationer for sorte i de to forskellige surveys. For at kunne sammenligne de to surveys, omkoder de forældrenes indkomstdata fra NLSY79 til at passe med NLS66. Forældrenes indkomstdata er opdelt i 11 grupper i NLS66 og samme gøres for NLSY79. Herefter bruger de medianen af indkomstgrupperne og bruger gennemsnittet af tre år. På trods af omkodningen er der stadig visse forskelle mellem de to surveys. I NLSY79 er det forældrene der rapporterer deres indkomst, mens det i NLS66 er børnene der rapporterer forældrenes indkomst. Den rapporterede forældreindkomst af børnene kan i nogle tilfælde valideres i et andet survey, eftersom nogle af fædre til børnene interviewes i National Longitudinal Survey of Older Men 1966. Derudover henviser de også til Levine & Mazumder (2002), som udregnede en reliability ratio på 0.93 for de to indkomstmål. Dette indikerer, at forskelle i de to mål ikke har den store betydning i forhold til bias af resultaterne. For børnene bruger de også et tre årigt gennemsnit af indkomsten. Forældrenes indkomst måles, når børnene er mellem 14 og 21 år, mens sønnernes indkomst måles, når de er i sluttiverne og sluttrediveerne. Dette vil derfor som udgangspunkt ikke have den store betydning i forhold til bias. Det kan dog være afgørende, hvor gamle forældrene er, når deres indkomst måles. Især når det anvendte data for indkomst, kun er et gennemsnit på tre år. Hvis ikke indkomsten måles, når den er tættest på den permanente indkomst, kan det skabe bias. Med dette i mente finder de statistisk signifikante resultater for IGE med relativt små standardfejl. Standardfejlene er omkring dobbelt så store for gruppen af sorte i forhold til gruppen af hvide. Dette kan muligvis skyldes de relativt færre observationer der er tilgængeligt for sorte.

Det sidste survey som bruges ofte i litteraturen fra USA er GSS. GSS er tværsnitsdata i modsætningen til PSID og NLS. Undersøgelsen blev første gang lavet i 1972 og er siden blevet indsamlet regelmæssigt af National Opinion Research Center hos University of Chicago. Levine & Mazumder (2002) bruger datasættet til at undersøge trenden i IGE. Deres indkomstdata for sønnerne kommer fra personer i alderen 28-36 år og deres resultater burde derfor umiddelbart ikke være påvirket af bias på baggrund af dette. Forældrenes indkomst måles, når deres søn er 16 år gammel. Eftersom GSS er tværsnitsdata indeholder det kun indkomstdata for det samme individ for ét år. Dette vil i høj grad være med til at skabe bias i estimatet og vil medføre en underestimering af IGE. Derudover er der stor usikkerhed omkring metoden, som bruges til at indsamle indkomstdata for forældrene. I GSS bliver børnene spurgt, når de er 16, hvordan de vil rangere deres families indkomst i forhold til den gennemsnitlige amerikanske familie. Dette gøres på en skala fra 1 til 5, hvor 1 er langt under gennemsnittet, 3 er gennemsnittet og 5 er langt over gennemsnittet. Ud fra disse oplysninger bruger Levine & Mazumder (2002) historisk data for indkomsten til at måle IGE. Der er umiddelbart stor tvivl

om, hvorvidt disse oplysninger er korrekte og kan være med til at skabe bias. I deres regressionsmodel for IGE har de 698 og 485 observationer. Når de bruger GSS opnås der ikke nogle statistisk signifikante resultater for IGE og deres standardfejl er relativt store.

Levine & Mazumder (2002) bruger udover GSS også NLS og PSID til at undersøge trenden i IGE i et forsøg på at sammenligne resultaterne fra de tre surveys. I den forbindelse beskriver de en del forskelle i mellem de tre surveys, som gør det vanskeligt at sammenligne resultaterne. Måden hvorpå indkomstdata indsamles i de tre surveys er forskelligt. Derudover er tidspunktet som indkomstdata indsamles på forskellig i de tre surveys. Dette gælder både for hvilket årstal der måles for indkomst og hvilken alder børnene har, når indkomsten måles. Mayer & Lopoo (2005) skriver følgende om PSID i forhold til de andre anvendte surveys (NLS og GSS): *"The PSID measure of parental income is likely to be more consistent than the measure from the NLS, the GSS or other survey data sets because the PSID is the only longitudinal data set that has asked relatively consistent questions about income"*. Til sidst beskriver Levine & Mazumder (2002) også bestemte karakteristiske forskelle i de tre surveys. I PSID er der en tendens til at oversample fattigere husholdninger. NLS oversampler sorte i både NLS66 og NLSY79. GSS er et survey, som repræsenterer hele landet.

I Storbritannien er det ofte The National Child Development Study (NCDS) og The British Cohort Study (BCS), som bruges til at undersøge den intergenerationelle mobilitet. Blanden et al. (2001) undersøger IGE i de to britiske surveys NCDS og BCS. NCDS er et survey af alle børn født i Storbritannien mellem den 3. og 9. marts 1958, mens BCS er et survey af alle børn født mellem den 5. og 11. april 1970. I NCDS spørges kohorten, når de er 7, 11, 16, 23, 33 og 42 år gamle, mens de i BCS spørges, når de er 5, 10, 16, 26 og 30 år gamle. I Blanden et al. (2001) måler de indkomsten, når børnene er 30 og 33 år gamle, hvilket betyder, at disse målinger vil være relativt tætte på deres permanente indkomst. Forældrenes indkomst måles, når børnene er 16 år gamle og forældrene vil derfor have forskellige aldre. Dette korrigeres i analysen. NCDS indeholder indkomstdata fra faderen og moderen, mens BCS indeholder indkomstdata for den samlede familie. Denne forskel gør, at det umiddelbart er svært at sammenligne resultater mellem de to surveys. I både NCDS og BCS bruger Blanden et al. (2001) ét års indkomstdata, hvilket betyder, at resultaterne højst sandsynligt lider af bias. Potentielt set kunne de også have brugt data fra NCDS, når kohorten var 42 år og BCS når kohorten var 26 år gamle. Dette vil have givet dem et års mere indkomstdata, men vil på samme tid også påvirke, hvilken alder deres indkomstdata kommer fra. I deres regressionsmodel har de omkring 2800 observationer for NCDS og 2000 for BCS. De finder signifikante resultater for IGE med relative små standardfejl.

Et af de store problemer i forhold til at undersøge den intergenerationelle mobilitet er tilgængelighed af data. I mange af de udviklede vestlige lande findes der som regel surveys, som minder om de beskrevne surveys ovenfor. Det samme gør sig ikke gældende for udviklingslande, hvilket gør det meget svært at undersøge mobiliteten i disse lande. I en stor ambitiøs udgivelse undersøger Narayan et al. (2018) den intergenerationelle mobilitet for 75 lande i verden. Overordnet set finder de, at der generelt er en lavere IGE i udviklede lande i forhold til udviklingslande. Det skal dog nævnes, at deres udgivelse som udgangspunkt fokuserer på uddannelsesmobilitet og at der er stor måleusikkerhed ved deres resultater for IGE. Den store måleusikkerhed skyldes metoden, som er brugt til at få indkomstdata i udviklingslandene. Disse lande har som sagt sjældent samme surveys, som er tilgængelige i udviklede lande, hvilket betyder, at indkomsten for forældrene er estimeret på baggrund af uddannelse. Dette gøres med en økonometrisk metode, som hedder two-sample two-stage least squares (TSTSLS). Kort fortalt estimerer metoden indkomsten ved tre trin. Først estimeres en indkomstfunktion for forældrene fra en ældre stikprøve, som er repræsentativ af den nuværende stikprøve. Herefter bruges de estimerede koefficienter af afkast til uddannelse og erfaring til at estimere forældrenes indkomst ved referencealderen. Dette gøres ved at bruge forældrenes alder og uddannelse fra den nuværende stikprøve. Til sidst laver man en regression af børnenes indkomst ved referencealderen på den estimerede forældreindkomst ved referencealderen (Björklund & Jäntti 1997). På denne måde er det altså muligt at estimere indkomsten for forældrene, selvom det ikke er tilgængeligt fra et bestemt survey. Med metoden følger, dog også nogle stærke antagelser som gør, at der er stor måleusikkerhed ved estimererne af den intergenerationelle mobilitet. Som udgangspunkt er det en stærk antagelse, at man kan estimere indkomsten ud fra en anden stikprøve. Det er usandsynligt, at individerne i den nuværende stikprøve vil have den samme mængde indkomst givet deres uddannelse og alder i forhold til den udvalgte stikprøve til at finde indkomstfunktionen. Denne usikkerhed er især stor, hvis der kun bruges et survey til at estimere ud fra. Med andre ord er der altså mange ting, som kan påvirke estimererne, som gør, at der er en stor måleusikkerhed forbundet med resultaterne.

For at opsummere findes der en del forskellige surveys i USA og Storbritannien. Overordnet set har de alle tilfælles, at der ikke er de store problemer med at få indkomstdata, når forældrene og børnene har den ønskede alder i forhold til at opnå så lidt bias som muligt. Det er som regel lidt sværere med data fra forældrene, eftersom de ofte vil have forskellige aldre, fordi de har fået børn på forskellige tidspunkter i deres liv. Et generelt problem som er tilstede i de forskellige surveys, er få år med indkomstdata fra både børn og forældre. GSS er tværsnitsdata og indeholder derfor kun ét år med ind-

komstdata. BCS indeholder som udgangspunkt også kun ét år med indkomstdata, mens NCDS indeholder to år med indkomstdata. PSID og NLS indeholder flere år med indkomstdata. For NLS bruges der et gennemsnit på tre år for forældre og børn, mens der for PSID bruges et gennemsnit på tre år for børnene og syv år for forældrene. Disse surveys indeholder altså nogle flere års data for indkomsten og vil derfor være bedre i forhold til at opnå lavere bias. I de fleste surveys er det individet selv, som oplyser om deres indkomst. Dette er ikke tilfældet i f.eks. GSS og NLS. I disse surveys bliver børnene spurgt, hvilken indkomst deres familie havde. Denne metode skaber stor usikkerhed og er højst sandsynligt med til at skabe bias. Disse ting er med til at understrege, at det er forskelligt, hvilken kvalitet de forskellige surveys har. Nogle egner sig bedre til at blive brugt til at undersøge den intergenerationelle mobilitet end andre. Generelt set viser litteraturen dog, at det er muligt at måle den intergenerationelle mobilitet med de forskellige surveys. Resultaterne vil dog ofte lide af en form for bias, eftersom der som regel er få års indkomstdata. Derudover er det også ofte et problem, at de forskellige surveys indeholder få observationer i forhold til registerdata. Det er få af de beskrevne undersøgelser, som indeholder mere end 2000 observationer i deres regressionsmodeller. Dette er markant mindre end undersøgelser, som bruger registerdata. Landersø & Heckman (2017) som bruger dansk registerdata har f.eks. 149.100 observationer og 621 observationer fra PSID. Få observationer skaber en del problemer i forhold til at undersøge den intergenerationelle mobilitet og kan være med til at skabe bias, fordi de kun beskriver en lille andel af den samlede population.

Et problem ved at undersøge den intergenerationelle mobilitet er, at der ikke altid er indkomstdata for forældre i de forskellige surveys. Dette problem kan løses ved at bruge den økonometriske metode TSTSTL. Narayan et al. (2018) bruger denne metode til at undersøge den intergenerationelle mobilitet i 75 lande. Stærke antagelser forbundet med metoden gør dog, at det kan være problematisk at bruge denne metode i praksis.

2.4. Data

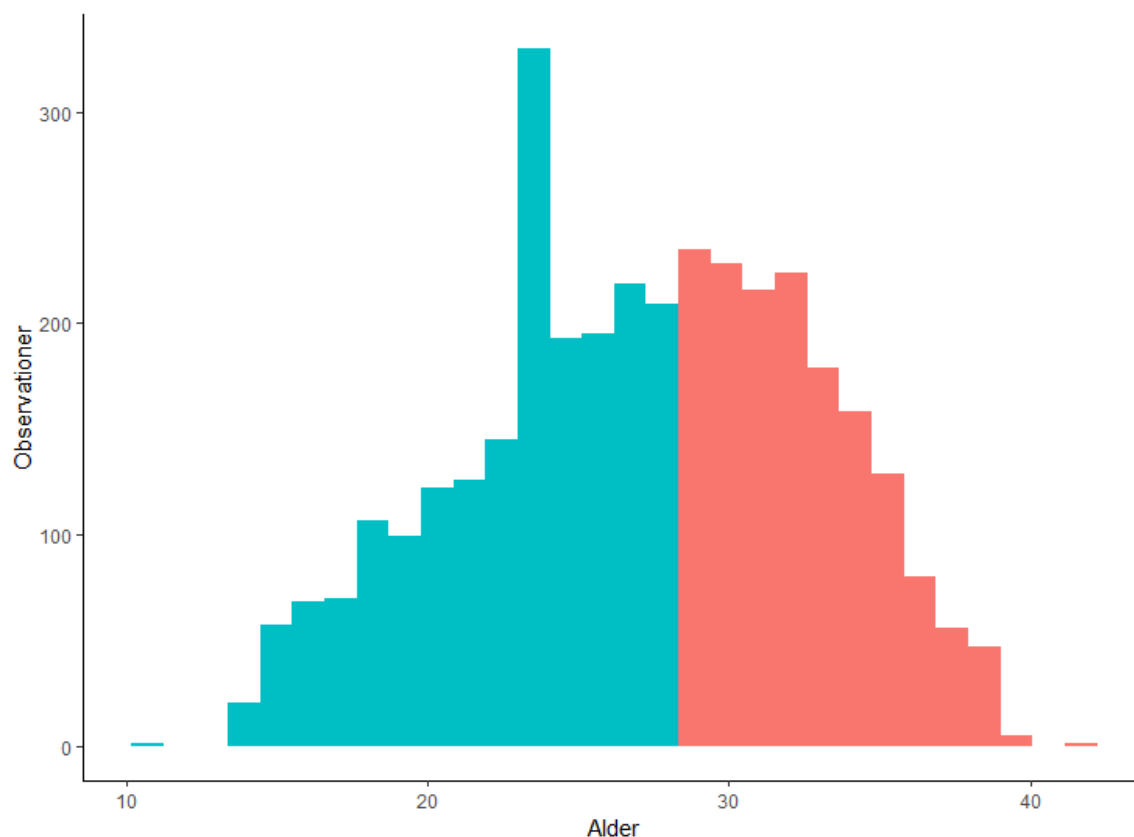
I dette afsnit introducerer vi det danske survey DLSY, som vi benytter os af til at illustrere de typiske udfordringer med survey datasættene beskrevet i forrige afsnit. Vi forklarer eventuelle tilskæringer, vi har lavet i forbindelse med analysen. Afsnittet fokuserer især på hvilke metodiske udfordringer, der opstår i forbindelse med at undersøge den intergenerationelle mobilitet med DLSY. Formålet med afsnittet er først og fremmest at præsentere det danske survey DLSY, men også i høj grad at vise de metodiske problematikker der opstår.

I projektet analyserer vi datasættet fra "Ungdomsforløbsundersøgelsen" også kendt som DLSY (Danish Longitudinal Survey of Youth). Dette er en undersøgelse, som følger en kohorte af 3151 danskere født i eller omkring 1954. Det første interview fandt sted i 1968, da deltagerne var omkring 14 år gamle og gik i 7. klasse. Herefter er deltagerne blevet interviewet i 1970, 1971, 1973, 1976, 1992, 2001 og 2004. Herudover er deltagernes forældre og klasselærer også blevet interviewet i 1969. Til sidst er deltagernes børn blevet interviewet i 2010 (DLSY-C) hvilket betyder, at datasættet dækker én familie igennem tre generationer. Surveyets struktur gør det til en rimelig repræsentant for klassen af surveys, vi er interesserede i mht. projektets problemformulering.

I forbindelse med tilskæringen af datasættet opstår der en del spørgsmål og problematikker i forhold til analysen af den intergenerationelle indkomstmobilitet. En af problematikkerne som er beskrevet i metodeafsnittet, omhandler betydningen af alder i forhold til børnenes og forældrenes indkomstdata. I vores tilfælde har forældrene en alder på omkring 38 år, når deres indkomst er målt. Dette stemmer umiddelbart godt overens med, hvad der anbefales i litteraturen. Det samme er ikke tilfældet, når det kommer til den tilgængelige indkomstdata for børnene. Eftersom dette er børn til kohorten, har de forskellige aldre. Det yngste barn i datasættet er 11 år og den ældste er 42 år. I forhold til mængden af observationer er dette problematisk. Hvis vi skulle være konsistente og måle indkomsten samme år, som vi måler indkomsten for forældrene, ville der kun være 41 observationer tilgængeligt. Vi tilskærer i stedet datasættet således, at børnene har forskellige aldre, for at få så mange observationer som muligt. Dette kan være problematisk i forhold til, at indkomsten bliver målt på forskellige tidspunkter af børnenes liv. Eftersom børnene har forskellige aldre, vil det være muligt, at nogle af børnene er et andet sted på deres indkomstkurve end forældrene. Hvis dette er tilfældet, vil det skabe et forkert billede af indkomsten og dermed skabe bias i estimatet.

I forhold til børnenes alder har vi valgt at følge Chetty et al. (2014), som påpeger, at life cycle bias stabiliserer sig omkring år 30. Landersø og Heckman (2017) anbefaler at måle indkomsten, når individerne er i sluttredive i Danmark. Vi kan dog ikke følge denne anbefaling uden at miste størstedelen af observationerne. Tilskæres dataet så vi har flest mulige observationer, består den samlede stikprøve af børn, som er 29 år og ældre. Vores baseline modeller vil derfor indeholde personer, som er 29 år og ældre, hvilket giver os 1153 observationer. I figur 2.2 kan der ses en illustration af tilskæringen af dataet. Observationerne markeret med blå er de observationer, som går tabt ved tilskæringen, mens de røde er observationerne i vores baseline model.

Figur 2.2 – Fordeling af observationerne for børnenes indkomst i forhold til alder.



Note: Figuren viser observationerne af børnenes indkomst for hele DLSY. Observationer markeret med blå er observationerne, som går tabt ved tilskæringen, mens de røde observationer er vores baseline.

Figur 2.2 viser, at størstedelen af observationerne i vores udvalgte data for børnenes indkomstdata ligger omkring sluttiverne og starttrediverne. Gennemsnitsalderen er omkring 32 år.

Som beskrevet i metodeafsnittet kræver det indkomstdata at kunne måle den intergenerationelle indkomstmobilitet. Dette er tilgængeligt fra forældrenes interview fra 1992 og 2001. Interviewet fra 2001 indeholder dog kun informationer om kohortens indkomst i intervaller, hvorimod de andre tilgængelige indkomstdata er i faktiske værdier. Vi benytter derfor kun indkomstdata fra deres interview i 1992.

Den tilgængelige indkomstdata for forældrene fra 1992 er angivet i forskellige værdier: månedsløn, løn pr. 14. dag, ugeløn og timeløn (før skat og fradrag). Datasættet indeholder informationer omkring, hvor mange timer de enkelte individer arbejder om ugen. Vi benytter denne information til at udregne de forskellige værdier for indkomst til en månedsløn (det antages, at fire uger er en måned). Dette betyder, at vi er i stand til at få sammenlignelige indkomstdata for alle forældrene. I vores analyse er disse individer specificeret som forældre, men indkomsten er dog opgjort således, at én observation

er enten en mor eller en far og derfor ikke samlet. Børnenes indkomstdata kommer fra de adspurgte børn af kohorten i et interview fra 2010. Her er den tilgængelige data angivet som månedlig bruttoindkomst i faktiske tal (før skat). Det er problematisk, eftersom den tilgængelige indkomstdata for forældrene er målt i lønindkomst. Bruttoindkomst er per definition større end lønindkomst, fordi den indeholder flere indtægter som f.eks. overførselsindkomster og kapitalindkomster. Vi har taget højde for inflation ved at skalere indkomsten i forhold til det danske forbrugerprisindeks (CPI). Deskriptiv statistik af indkomsten kan ses i tabel 2.1.

Tabel 2.1 – Deskriptiv statistik af indkomstdata.

	Gennemsnit	S.E.	1. kvartil	Median	3. kvartil	Maks	Observationer
Børn	334.300	260.385	246.000	312.000	384.000	4.776.000	1153
Forældre	283.928	119.590	212.716	265.895	319.074	1.772.631	1153

Note: Tabellen viser deskriptiv statistik af indkomstdata fra observationerne i vores baseline model.

Tabel 2.1 viser, at gennemsnitsindkomsten for børnene og forældrene ligger relativt tæt på hinanden. Børnene har en gennemsnitsindkomst på 334.300 kr., mens forældrene har en gennemsnitsindkomst på 283.928. Det er værd at ligge mærke til den store forskel i maksværdien af indkomsten mellem børnene og forældrene. I børnenes indkomstdata er der en maksværdi på 4.776.000 kr., mens den er på 1.772.631 kr. for forældrene. Derudover ses der også en relativ stor standardafvigelse i børnenes indkomstdata.

I metodeafsnittet beskrives der forskellige problematikker i forhold til at estimere den intergenerationale indkomstmobilitet med få års indkomstdata. Som udgangspunkt vil estimater baseret på få års indkomstdata have større bias end estimater baseret på flere års indkomstdata. Eftersom vores datasæt kun indeholder indkomstdata for ét år, vil dette begrænse os i forhold til at tage højde for denne bias.

2.5. Analyse

I det følgende afsnit vil vores analyse af den intergenerationale indkomstmobilitet i det danske datasæt DLSY blive præsenteret. Metoderne der bruges til analysen er IGE, rank-rank korrelation og transitionsmatrice, hvilke er beskrevet tidligere i metodeafsnittet. Validiteten af resultaterne diskuteres i forhold til kendte problematikker ved at måle den intergenerationale indkomstmobilitet i små surveys. Formålet med analyserne er at påvise, hvilke metodiske problematikker der opstår, når man undersøger den intergenerationale mobilitet med deskriptive mål i surveys.

2.5.1. Intergenerationel elasticitet

I dette afsnit vil resultaterne af vores analyse af den intergenerationelle indkomstelasticitet (IGE) i det danske datasæt DLSY blive vist. Først præsenteres der en undersøgelse fra litteraturen, for at få en fornemmelse af, hvordan estimaterne for IGE generelt ser ud. Herefter opstiller vi modellen ud fra metodeafsnittet og viser resultaterne af analysen. Analysen består af en baseline model og forskellige aldersgrupper. Derudover undersøger vi hvorvidt, det er muligt at opdele efter far og søn og far og datter samt lave en opdeling efter geografi. Dette gøres ofte i litteraturen, eftersom det er interessant at undersøge, hvorvidt der er forskellige resultater for de forskellige inddelinger. Som beskrevet tidligere opstår der generelt en del problematikker i forhold til at måle IGE i små surveys. Hvilken betydning disse problematikker har i forbindelse med vores analyse af IGE i DLSY diskuteres i løbet af afsnittet.

Den intergenerationelle indkomstmobilitet er interessant i sig selv, men ved at sammenligne med mobilitetsmål fra andre lande kan der opnås en bedre forståelse for mobiliteten. I litteraturen ses det ofte, at der bliver lavet komparative analyser af den intergenerationelle indkomstmobilitet. Eftersom det ikke er alle lande, der har den nødvendige data tilgængeligt, er det ofte komparative analyser mellem USA, Storbritannien og de nordiske lande. Desuden er disse lande interessante at undersøge i forhold til de vidt forskellige samfundsmodeller. Jäntti et al. (2006) har analyseret den intergenerationelle indkomstmobilitet for USA, Storbritannien, Danmark, Finland, Norge og Sverige. De analyserer National Longitudinal Survey of Youth (NSLY) i USA, National Child Development Study (NCDS) for Storbritannien, mens de bruger registerdata for de nordiske lande. For at kunne sammenligne resultaterne på tværs af landene, bruger de kun indkomstdata fra ét år. Dette er et godt eksempel på nogle af de problemer, som forskere kan have, når de forsøger at sige noget om mobiliteten på tværs af lande. For at begrænse life cycle bias bruger de fædre, som er 42 år gamle og børn der er mellem 30 og 42 år gamle. Deres estimater mellem fædre og sønner kan ses i tabel 2.2.

Tabel 2.2 – IGE-estimer fra nordiske lande, USA og Storbritannien.

Land	Intergenerationel elasticitet
Danmark	0.071
Finland	0.173
Norge	0.155
Sverige	0.258
Storbritannien	0.306
USA	0.517

Kilde: Jäntti et al. 2006 – Tabellen viser estimer af IGE for forskellige lande. Estimerne er baseret på indkomstdata fra et år.

Her ses der en tydelig forskel mellem USA, Storbritannien og de nordiske lande, hvilket også er det generelle resultat i litteraturen. Desuden ses det, at Danmark har den lavest elasticitet og dermed den højeste mobilitet.

Den intergenerationelle elasticitet bliver som beskrevet i metodeafsnittet estimeret ved en log-log regression mellem børnenes og forældrenes indkomst. Vi følger den generelle litteratur og opsætter en regression for IGE i ligning 2.23:

$$\log(Y_{børn}) = \alpha + \beta \log(Y_{forældre}) + ALDER + ALDER^2 + \varepsilon \quad (2.23)$$

Hvor, $Y_{børn}$ = indkomst børn, $Y_{forældre}$ = indkomst forældre, β = IGE

Eftersom børnene og forældrene har forskellige aldre, når vi måler indkomsten, inddrager vi alder som en kontrolvariabel for at korrigere for dette. Som beskrevet i metodeafsnittet, er der forskellige resultater i litteraturen i forhold til, hvilken alder det er mest optimalt at måle indkomsten. Vores baseline indeholder IGE, hvor børnene er 29 år eller ældre. Derudover har vi valgt at undersøge forskellige inddelinger af alder for at teste, hvorvidt det har nogen betydning for resultatet af IGE. Vi tester derfor også for IGE, hvor $ALDER = >30, >32$ og >33 . Resultaterne af undersøgelsen kan ses i tabel 2.3:

Tabel 2.3 – IGE-estimer i DLSY fra forskellige aldersopdelinger

	ALDER= >28	ALDER= >30	ALDER= >32	ALDER= >33
IGE	0.074* (0.035)	0.09* (0.041)	0.098* (0.047)	0.046 (0.054)
Observationer	1147	792	458	335

Note: Tabellen viser estimer af IGE for forskellige inddelinger af alder i DLSY.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.

Som det kan ses i tabel 2.3, giver vores baseline model et statistisk signifikant resultat af IGE på 0.074. Selvom vores analyse giver et statistisk signifikant resultat, kan vi som udgangspunkt ikke opnå et estimat af IGE uden en vis måleusikkerhed, eftersom børnene og forældrene har to forskellige opgørelser for indkomst i datasættet. Desuden er det værd at ligge mærke til, at standardfejlene for estimatet er relativt stor. Dette kan også ses illustreret i figur A.1 i appendiks A. Det vil altså sige, at estimatet indeholder stor måleusikkerhed og det kan derfor være svært at sige noget om det faktiske niveau for mobiliteten. På trods af at der er stor måleusikkerhed for vores estimat af IGE, stemmer resultatet godt overens med den eksisterende litteratur. Jäntti et al. (2006) finder en IGE på 0.071 mellem sønner og fædre. Det kan dog være svært at sammenligne disse resultater, eftersom vores metode inkluderer begge børn og forældre. Landersø & Heckman (2017) inkluderer ligesom os begge børn og forældre og minder derfor om vores fremgangsmåde. De finder forskellige estimater af IGE alt efter, hvilken form for indkomst der undersøges. Deres estimat af IGE for lønindkomst er på 0.083. Ligheden mellem vores resultat og litteraturen er mærkværdig, men kan ikke forklares ud fra et statistisk grundlag. Det er dog værd at notere, at både Jäntti et al. (2006) og Landersø & Heckman (2017) baserer deres analyse på registerdata fra Danmark, hvilket betyder, at de har markant flere observationer. De bruger en stikprøve på henholdsvis 60.000 og 150.000 observationer til deres analyse, hvilket også betyder at de har betydeligt mindre standardfejl og dermed estimater med mindre måleusikkerhed.

Generelt set kan koefficienten bruges til at sammenligne med koefficienter fra andre lande. Landersø & Heckman (2017) finder f.eks. en IGE på 0.289 for lønindkomst i USA. Her er der en tydelig forskel mellem den danske og amerikanske IGE, hvor der relativt set er højere intergeneretionel indkomstmobilitet i Danmark. Derudover kan koefficienten også bruges isoleret set til at se effekten af en stigning i forældrenes indkomst. Ud fra vores resultat (0.074) vil en forskel i indkomsten på 10 procent hos to forskellige forældre betyde en forskel på 0.74 procent mellem deres børns indkomst. Det vil altså sige, at forskelle i forældres indkomst ikke vil have stor betydning for børnenes fremtidige indkomst.

I metodeafsnittet er der beskrevet forskellige problematikker i forhold til at undersøge IGE. En af problematikkerne der specifikt opstår i forhold til at måle IGE, er at det ikke er muligt at inddrage personer, som har en indkomst på nul. Eftersom der findes personer, som ikke har en indkomst af forskellige årsager, kan det gøre målet mindre realistisk. I vores datasæt er der seks personer, der udgår af vores analyse, fordi de har en indkomst på nul. Det er relativt få og ses derfor ikke som et større problem.

En anden problematik som umiddelbart kan have haft en betydning på resultatet af estimatet, er hvilke år, der måles indkomstdata for børnene og forældrene. Som beskrevet i dataafsnittet, er forældrene omkring 38 år, når deres indkomstdata bliver målt. Dette passer godt overens med størstedelen af litteraturen, som påpeger, at det er optimalt at måle forældrenes/farens indkomstdata fra omkring midten af trediverne til midten af fyrrerne (Mazumder, 2005; Baker & Solon, 2003). Lidt anderledes passer det dog med børnenes indkomstdata. Denne er som vist i figur 2.2 målt for børn med forskellige aldre. Problematikken ved dette er som nævnt tidligere, at børnene potentielt set kan være forskellige steder på deres indkomstkurve. Hvis dette er tilfældet, vil det give inkonsistente målinger af indkomsten. En person på 28 år har f.eks. andre forudsætninger i forhold til indkomst sammenlignet med en person på 40 år. Derudover skaber det også inkonsistente målinger i forhold til forældrenes indkomst, som er målt ved den samme alder. Hvis ikke indkomsten er målt omkring den samme alder hos både børn og forældre, kan det potentielt set påvirke resultatet, fordi de har forskellige forudsætninger alt efter deres alder. I relation til dette kan der også opstå life cycle bias, fordi indkomsten er målt ved en relativ ung alder for børnene i vores analyse. Life cycle bias kan som beskrevet i metodeafsnittet have indflydelse på IGE, hvis indkomsten måles på et tidspunkt, hvor individernes indkomst endnu ikke er ”modnet”. Med andre ord handler det om at måle indkomsten på et tidspunkt, hvor forskellen mellem personer med forskellige indkomstudviklinger passer med forskellen på deres langsigtede indkomstkurver. I litteraturen findes der forskellige bud på, hvornår det mest optimale tidspunkt er. Generelt set anbefales det at måle fra omkring midt trediverne til midt fyrrerne, men som nævnt tidligere i afsnittet viser ny forskning fra Chetty et al. (2014), at det allerede er muligt at måle indkomsten omkring 30 år. Landersø & Heckman (2017) viser dog, at det i Danmark er bedst at måle indkomsten i slutningen af trediverne. Denne forskning er baseret på administrative datasæt, hvilket skaber en vis usikkerhed om, hvorvidt det samme er tilfældet for små surveys. Vi kan derfor ikke udelukke måleusikkerhed i forhold til at inddrage børns indkomst fra de er 29 år. I et forsøg på at undersøge indflydelsen af dette, har vi undersøgt IGE for forskellige aldersgrupperinger. Som vist i tabel 2.3 er IGE yderligere undersøgt for aldersgrupper, hvor børnene er over 30 år, over 32 år og over 33 år. Ud fra resultaterne kan det ses, at IGE stiger på samme tid med, at alderen stiger. Dette kan være med til at give en indikation af, at vores baseline model lider af life cycle bias. Vi kan dog som udgangspunkt ikke konkludere dette, eftersom der kan være andre ting end alder, som har påvirket individernes indkomst. Derudover kan det også ses, at standardfejlene stiger sammen med alderen, hvilket højt sandsynligt skyldes, at mængden af observationer falder på samme tid. De stigende standardfejl gør, at vi som udgangspunkt ikke kan sige noget om, hvorvidt der er life cycle bias. Størrelsen

på standardfejlene gør, at estimaterne mere eller mindre ligger inde for de samme intervaller, hvilket betyder, at det ikke har den store betydning at ændre aldersgrupperne. En illustration af måleusikkerheden ved resultaterne kan ses i figur A.1 i appendiks A. Desuden ses det også, at estimatet mister sin signifikans, når alderen er større end 33 år.

Overordnet set må det antages, at vores resultater af IGE lider af attenuation bias. Denne bias opstår som beskrevet i metodeafsnittet, når IGE bliver estimeret med få års indkomstdata. Generelt set falder attenuation bias, når der bruges indkomstdata fra flere år. Vores datasæt indeholder kun indkomstdata for ét år og vil derfor højst sandsynligt være påvirket af bias. Hvis estimatet lider af attenuation bias, vil der være en relativ lav attenuation faktor. Mazumder (2005) viser ved simulationer, at der ved brug af et års indkomstdata vil være en attenuation faktor på 0.5. Det vil altså sige, at IGE vil blive underestimeret med omkring 50%. Det er svært at vurdere hvor meget vores estimat er underestimeret, men det må antages at estimatet er underestimeret. Hvis det antages, at vores estimat er underestimeret med 50%, kan det korrigeres for ved at gange med 2 og dermed opnå et estimat på 0.148.

Vi har som udgangspunkt undersøgt IGE, hvor vi inkluderer både mødre og fædre og sønner og døtre. I litteraturen undersøges IGE ofte mellem far og søn eller far og datter. Denne opdeling kan umiddelbart være problematisk, hvis datasættet indeholder få observationer. Vi har undersøgt hvorvidt det er muligt med det danske datasæt DLSY. Resultatet af regressionerne ses i tabel 2.4.

Tabel 2.4 – IGE-estimer i DLSY fra opdelingen af far og søn og far og datter

	Far og søn	Far og datter
IGE	0.11 (0.07)	0.068 (0.06)
Observationer	349	401

Note: Tabellen viser estimer af IGE for opdelingen efter far og søn og far og datter.

**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.*

Som det kan ses i tabel 2.4, finder vi ikke nogen statistiske signifikante resultater, når vi inddeler i grupper med far og søn og far og datter. Ydermere er standardfejlene for estimaterne relativt høje, hvilket betyder, at det er svært at sige noget om mobiliteten. Dette illustreres i figur A.2 i appendiks A. Det er også svært at bedømme, om der er nogen forskel mellem de to inddelinger pga. de store standardfejl. Dette skyldes højst sandsynligt de få observationer, som er tilgængeligt i disse opdelinger. Vi kan derfor som udgangspunkt ikke sige noget om IGE, når vi undersøger for far og søn og far og datter.

En anden retning indenfor litteraturen undersøger, hvorvidt geografi har nogen betydning for størrelsen af IGE. Chetty et al. (2014) undersøger f.eks., om der er forskellige mål af IGE, hvis det undersøges i forskellige områder i USA. Denne form for undersøgelse kan for så vidt godt laves i det danske datasæt DLSY, men mængden af observationer for hver kommune, gør det svært at få nogle resultater, som er signifikante. Datasættet indeholder informationer for forældrene fra 136 kommuner² og det højeste antal observationer fra én kommune er 51. De fleste kommuner har mindre end 10 observationer, hvilket gør det svært at få meningsfulde resultater. En anden opdeling, der er muligt at lave i datasættet, er efter hvorvidt forældrene bor i København, en forstad til København, en stor by, en lille by, en landsby eller et landdistrikt. Resultaterne af analysen kan ses i tabel 2.5.

Tabel 2.5 – IGE-estimerer fordelt på størrelse af byer.

	KBH	Forstader KBH	Stor by	Lille by	Landsby	Landdistrikt
IGE	0.30• (0.16)	0.13 (0.13)	0.05 (0.07)	0.12 (0.8)	-0.15 (0.7)	0.05 (0.09)
Observationer	66	104	325	240	236	176

Note: Tabellen viser estimerne af IGE for forskellige størrelser af byer. Definitioner af byer: Stor by: >10.000 indbyggere. Lille by: 2.000-10.000 indbyggere. Landsby: 200-2.000 indbyggere.

**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.*

Som det kan ses i tabel 2.5, finder vi ikke nogen statistisk signifikante resultater for de forskellige opdelinger, bortset fra individerne der bor i København. Dette er overraskende, eftersom dette er opdelingen, som indeholder færrest observationer, men skyldes højst sandsynligt det overraskende store estimat. Det er desuden værd at ligge mærke til, at standardfejlene igen er meget store, hvilket betyder, at der er stor måleusikkerhed forbundet med estimerne. Dette er højst sandsynligt pga. de få observationer tilgængeligt. En illustration af måleusikkerheden kan ses i figur A.3 i appendiks A.

For at opsummere findes der en del usikkerheder og problematikker i forhold til vores analyse af IGE i det danske datasæt DLSY. Disse kan i høj grad overføres til generelle problemer ved at undersøge IGE i små surveys. I vores tilfælde er indkomstdata opgjort forskellige for børnene og forældrene. Det kan være en generel problematik ved surveys, at de ikke er konsistente i den måde de spørger individerne, hvilket vil gøre det svært at sammenligne to grupper. Derudover kan der også opstå usikkerheder i forhold til, hvilken alder børnene og forældrene har, når deres indkomst er målt. Generelt kan det være et problem at få indkomstdata fra de ”optimale” år i surveys. Dette er også et problem i

² Dette er før sammenlægningen af kommunerne, som i dag kun består af 98 kommuner.

vores datasæt, som kun indeholder en begrænset mængde data mellem midt tredive og midt fyrrerne (nyere forskning indikerer, at dette måske ikke er nødvendigt). Surveys med få observationer er et gennemgående problem med at undersøge IGE og er også et problem i det danske datasæt DLSY. Konsekvensen af dette er at, vi er nødt til at bruge indkomstdata for børnene, hvor de har forskellige aldre. Dette er med til at skabe inkonsistente målinger af indkomsten og dermed usikkerhed i estimatet for IGE. Et sidste problem som generelt er tilstede, når IGE undersøges ved hjælp af surveys, er få års indkomstdata. Dette er med til at skabe attenuation bias og dermed underestimere IGE. Derudover viser det sig også, at det kan være svært i små surveys at undersøge forskellige opdelinger af datasættet. Få observationer i det danske datasæt DLSY gør det ikke muligt at undersøge far og søn og far og datter meningsfyldt. Derudover er det heller ikke muligt at lave geografiske opdelinger efter kommune, som er udbredt i litteraturen. Det var derimod muligt at opdele efter forskellige størrelser af byer. Disse problematikker kan i høj grad være til stede i de fleste små surveys, eftersom de ofte er præget af få observationer.

2.5.2. Rank-rank korrelation

I dette afsnit præsenteres vores analyse af rank-rank korrelationen i det danske datasæt DLSY. Først præsenteres modellen. Derefter vises resultaterne af analysen med vores baseline model og disse sammenlignes med resultater fra litteraturen. Herudover forsøger vi ligesom med IGE at opdele datasættet efter far og søn og far og datter samt geografi. Til sidst diskuteres eventuelle metodiske problematikker i forhold til analysen og de generelle problemer i forhold til at undersøge rank-rank korrelation i små surveys.

Rank-rank korrelation er som beskrevet i metodeafsnittet en regression af børnene og forældrenes indkomstdata opdelt i percentiler. Ligningen for regressionen ses i ligning 2.24.

$$R_{børn} = \alpha + \beta_{PR} P_{forældre} + \varepsilon \quad (2.24)$$

Hvor, $R_{børn}$ = indkomstgruppe i for børnene, $P_{forældre}$ = indkomstgruppe i for forældrene, β_{PR} = rank – rank korrelation

Ligesom i analysen for IGE indeholder vores baseline indkomstdata for forældre, som er omkring 38 år og indkomstdata for børnene som er ældre end 28 år. Derudover kontrolleres der også for alder. Dette gøres ved at lave to regressioner af børnene og forældrenes indkomst på alder og alder². Herfra bruges residualerne fra regressionen til at lave ranken for indkomsterne. Denne metode blev påvist af Frisch & Waugh (1933). Resultaterne af regressionen kan ses i tabel 2.6.

Tabel 2.6 – Rank-rank korrelation for baseline model.

Rank-rank korrelation	0.049• (0.029)
Observationer	1153

Note: Tabellen viser rank-rank korrelationen for vores baseline model i DLSY.

**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.*

En af fordelene ved at foretage en analyse af rank-rank korrelationen, er at vi kan inddrage personer, som har en indkomst på nul. Dette forklarer, at vi i denne analyse har flere observationer end i analysen af IGE. Som det kan ses i tabel 2.6, finder vi en statistisk signifikant rank-rank korrelation på 0.049 i det danske datasæt DLSY. Som beskrevet ved IGE er der stor usikkerhed omkring resultatet på trods af et statistisk signifikant resultat, fordi vi benytter to forskellige opgørelser for indkomst. Desuden finder vi ligesom med IGE relative store standardfejl, som gør det svært at sige noget om mobiliteten. En illustration af dette ses i figur A.4 i appendiks A. Det er heller ikke muligt at finde noget lignende resultat, hvis vi sammenligner med den eksisterende litteratur. Boserup, Kopczuk, og Kreiner (2013) finder en rank-rank korrelation på 0.18 for Danmark med formuedata og Landersø & Heckman (2017) finder et estimat på 0.205 for lønindkomst i Danmark. Begge udgivelser bruger administrative data. Hvis det antages, at vi kunne sige noget ud fra vores resultat, ville vi kunne aflæse, at en stigning i forældrenes rank på 10 percentilpoint ville medføre en stigning på 0.49 percentilpoint i børnenes rank.

Hvorfor vores resultat ikke passer med den eksisterende litteratur, kan være svært at specificere, men kan forklares med flere ting. Først og fremmest kan det forklares med, at vi som sagt ikke bruger den samme opgørelse for indkomst mellem børnene og forældrene. Derudover kan der som beskrevet i metodeafsnittet også opstå problemer med attenuation bias og life cycle bias, som vil medføre, at målet vil blive underestimeret. Ligesom med IGE anbefales det at bruge et gennemsnit over flere år for indkomsten. Dette er ikke muligt i vores datasæt og det må derfor antages, at estimatet lider af attenuation bias og er underestimeret af denne grund. Mazumder (2005) påviste, at der potentielt kunne forefindes et attenuation bias på 0.69. Vores punkt-estimat (uagtet standard-fejlen for estimatet) udgør her kun omkring $\frac{1}{4}$ af estimerne opnået med det administrative data, hvilket svarer til en attenuation faktor på 0.25. En vis usikkerhed opstår også i relation til life cycle bias. Det anbefales som udgangspunkt ligesom med IGE at måle indkomsten, når individerne er tættest på deres permanente indkomst. Som beskrevet tidligere, er der forskellige bud på hvornår indkomsten kan måles og eftersom vi har inddraget individer, som er relativt unge, kan det ikke udelukkes, at vores estimat er påvirket af life cycle bias. Problemet med life cycle bias anses dog ikke for at være et mindre problem

i rank-rank korrelation end ved IGE. Ydermere, er det ligesom ved IGE problematisk, at vores indkomstdata for børnene er spredt ud på flere forskellige aldre. I metodeafsnittet beskrives rank-rank korrelationen som at være tættere på linearitet end IGE. Det er derfor muligt at målet indeholder mindre måleusikkerhed end IGE. Dette skyldes, at der i litteraturen er fundet tegn på ikke-linearitet i estimater for IGE mellem fædre og sønner i Skandinavien (Bratsberg et al. 2007). Det er derfor også muligt, at det samme er gældende for vores estimater af IGE. Som udgangspunkt burde det betyde, at estimaterne for rank-rank korrelationerne er mere korrekte, eftersom de har et bedre lineært ”fit”.

Ligesom ved undersøgelsen af IGE har vi undersøgt, hvorvidt det er muligt at opdele datasættet således, at vi undersøger rank-rank korrelationen for far og søn og far og datter. Resultaterne af undersøgelsen ses i tabel 2.7.

Tabel 2.7 – Rank-rank korrelationer for opdelingen af far og søn og far og datter.

	Far og søn	Far og datter
Rank-rank korrelation	0.092• (0.053)	0.12* (0.05)
Observationer	351	403

Note: Tabellen viser rank-rank korrelationerne for opdelingen efter far og søn og far og datter i DLSY.

**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.*

Som det kan ses i tabel 2.7, giver undersøgelsen af rank-rank korrelationer for far og søn og far og datter statistiske signifikante resultater. Der er ligesom ved tidligere estimater stor måleusikkerhed ved resultaterne pga. forskellige indkomstopgørelser og relative store standardfejl. Måleusikkerheden ved standardfejlene kan ses i figur A.5 i appendiks A. Det er dog muligt at sammenligne resultaterne med resultaterne uden opdelingen. Her ses der en ret markant forskel i rank-rank korrelationen. Resultaterne med opdelingerne er mere eller mindre fordoblet i forhold til resultaterne med det samlede datasæt, som vist i tabel 2.7. Det tyder altså på, at hvis man udelukkende ser på resultaterne for far og søn og far og datter, vil man højst sandsynligt overestimere rank-rank korrelationen og dermed underestimere mobiliteten. Derudover ses der en lille forskel mellem korrelationen for far og søn og far og datter.

Vi undersøger ligesom for IGE, hvorvidt det er muligt at lave geografiske opdelinger i undersøgelsen af rank-rank korrelation. Vi kan ikke opdele efter kommuner pga. få observationer og opdeler derfor efter størrelser på byer. Resultaterne af undersøgelsen ses i tabel 2.8.

Tabel 2.8 – Rank-rank korrelationer for opdelinger efter størrelse af by.

	KBH	Forstader KBH	Stor by	Lille by	Landsby	Landdistrikt
Rank-rank korrelation	0.03 (0.124)	0.029 (0.099)	0.02 (0.554)	0.11• (0.064)	-0.01 (0.065)	0.09 (0.076)
Observationer	67	104	328	242	236	176

Note: Tabellen viser estimerne af rank-rank korrelationer for forskellige størrelser af byer. Definitioner af byer: Stor by: >10.000 indbyggere. Lille by: 2.000-10.000 indbyggere. Landsby: 200-2.000 indbyggere.

**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$.*

Som det kan ses i tabel 2.8, finder vi ikke nogen statistisk signifikante resultater for rank-rank korrelationerne bortset fra opdelingen, hvor forældrene kommer fra en lille by. Det statistisk signifikante resultat minder om resultaterne for opdelinger af far og søn og far og datter. Derudover er det igen værd at ligge mærke til de store standardfejl, som medfører en betydelig måleusikkerhed omkring estimerne. Dette illustreres i figur A.6 i appendiks A.

Generelt set findes der lignende problematikker i forhold til at undersøge rank-rank korrelation i små surveys som ved undersøgelsen af IGE. I vores tilfælde er det svært at sige noget ud fra vores estimer, eftersom det danske datasæt DLSY indeholder to forskellige opgørelser for indkomst. Ligesom ved IGE er der generelle problemer med attenuation bias og life cycle bias, når rank-rank korrelationen undersøges. Problemet med life cycle bias anses dog for at være et mindre problem i estimeringen af rank-rank korrelationen end ved IGE. Vores estimat af rank-rank korrelationen for baseline er relativt langt fra, hvad der findes i litteraturen af Chetty et al. (2014) og Landersø & Heckman (2017), som finder en rank-rank korrelation på henholdsvis 0.18 og 0.205. Dette kan muligvis forklare, at vores resultater kun indeholder ét år med indkomstdata og derfor er påvirket af bias. Det vil forklare, at vi underestimerer rank-rank korrelation og målet derfor i virkeligheden burde være højere. Til sidst undersøger vi, hvorvidt det er muligt at måle rank-rank korrelationen, når der opdeles efter far og søn og far og datter samt geografi. Her finder vi statistisk signifikante resultater for opdelingen af far og søn og far og datter, hvilket betyder, at det er muligt at undersøge dette i det danske datasæt DLSY. Resultaterne er dog påvirket af stor måleusikkerhed, hvilket kan ses ved det relativt store standardfejl. Når vi opdeler datasættet efter geografi, finder vi ikke nogen statistisk signifikante resultater med undtagelse af opdelingen med lille byer. Vi finder ligesom ved de andre resultater store standardfejl, hvilket indikerer stor måleusikkerhed ved målene.

2.5.3. Transitionsmatrice

I dette afsnit præsenteres vores resultater af analysen for transitionsmatricer i det danske datasæt DLSY. Først forklares opsætningen af matricen. Herefter vises resultaterne. Til sidst diskuteres eventuelle problemer med resultaterne.

Transitionsmatricerne er lavet ud fra den samme opdeling, som benyttes i rank-rank korrelationen. Indkomsten for børnene og forældrene er derfor fordelt i et interval mellem 0 og 1. Herefter opdeler vi individerne i kvintiler som gør, at vi har fem indkomstgrupper for børnene og fem indkomstgrupper for forældrene. Til sidst laves der en sandsynlighedsmatrice, hvor resultaterne kan ses i figur 2.3.

Figur 2.3 – Transitionsmatrice for observationerne i vores baseline model.

		Kvintil - forældre				
		1	2	3	4	5
Kvintil - børn	1	0,19	0,22	0,21	0,19	0,20
	2	0,26	0,19	0,22	0,17	0,16
	3	0,17	0,19	0,18	0,23	0,23
	4	0,21	0,24	0,20	0,19	0,16
	5	0,17	0,16	0,20	0,23	0,24

Note: Figuren viser sandsynligheden for at børnene ender i samme indkomstgruppe som deres forældre i vores baseline model fra DLSY. Antal observationer: 1153

Som det kan ses ud fra figur 2.3, er der umiddelbart ingen sammenhæng mellem hvilken indkomstgruppe børnene ender i ud fra deres forældres indkomstgruppe. Hvis et barn har en forælder der kommer fra 1. kvintil, vil der være 19% sandsynlighed for, at barnet selv vil være i 1. kvintil og 17% sandsynlighed for at barnet vil være i 5. kvintil. Overraskende viser resultaterne altså, at det mere eller mindre er tilfældigt, hvilken indkomstgruppe man ender i. Generelt må det antages, at der som udgangspunkt vil være større sandsynlighed for, at børnene ender i samme indkomstgruppe som deres forældre. Dette ses i enkelte tilfælde. Et barn hvis forælder kommer fra 5. kvintil, vil f.eks. have størst sandsynlighed for at ende i samme indkomstgruppe. Resultatet er også overraskende med tanke på, at vores indkomstdata for børnene er målt som bruttoindkomst. Som udgangspunkt ville det forventes,

at der ville være en større sammenhæng i de højere indkomstgrupper. En forventning om dette skyldes, at bruttoindkomsten bl.a. medregner kapitalindkomster, hvilket individer fra en højere indkomstgruppe burde have en større andel af. Det er umiddelbart svært at forklare resultaterne, men som beskrevet i metodeafsnittet, kan lignende bias også opstå ved transitionsmatricer, som ved undersøgelser af IGE og rank-rank korrelation. Ifølge litteraturen vil der opstå en overestimering af mobiliteten (især i de nederste indkomstgrupper), hvis der benyttes få års indkomstdata. Eftersom dette er tilfældet i vores undersøgelse, må det antages, at vores resultat overestimerer mobiliteten. Dette vil være med til at forklare den overraskende høje mobilitet, vi finder i transitionsmatricen.

Vi har ligesom ved undersøgelsen af IGE og rank-rank korrelationer opdelt datasættet i far og søn og far og datter for at undersøge, hvorvidt det er muligt at lave transitionsmatricer med disse opdelinger. I modsætningen til undersøgelsen af IGE og rank-rank korrelationer er der ikke nogen måling af, hvorvidt resultaterne er statistisk signifikante. Vi kan derfor godt lave transitionsmatricer for relativt små dataopdelinger. Der vil dog altid være en vis måleusikkerhed, hvis datasættet indeholder relativt få observationer. Transitionsmatricerne for opdelingen af far og søn og far og datter kan ses i figur 2.4. Antallet af observationer er angivet under figuren.

Figur 2.4 – Transitionsmatricer for far og søn og far og datter.



Note: Figuren viser sandsynligheden for at sønner og døtre ender i samme indkomstgruppe som deres far i DLSY. Antallet af observationer: Far og søn = 351 og Far og datter = 403

I modsætningen til transitionsmatricen for det samlede datasæt, viser der sig at være en relativt større tendens i transitionsmatricerne, når vi opdeler for far og søn og far og datter. Figur 2.4 viser, at der er en forholdsvis lav mobilitet i bunden og i toppen for opdelingen af far og søn. Det vil altså sige, at

en søn hvis far er i den laveste indkomstgruppe, også højst sandsynligt vil ende i den nederste indkomstgruppe. Det samme gør sig gældende for den øverste indkomstgruppe. Derudover ses det også, at der generelt er en tendens til, at sønnerne ender i den samme indkomstgruppe, som deres far. Tendensen er dog mindre klar end ved bunden og toppen. Transitionsmatricen for far og datter viser en relativ mindre tendens med undtagelse for toppen. Hvis en datter har en far, som kommer fra den øverste indkomstgruppe, er der også størst sandsynlighed for, at hun selv ender i den øverste indkomstgruppe. Det samme gør sig gældende for 4. kvintil. Herefter bliver tendensen mere uklar. Noget tyder på, at der er en tendens til, at døtrene i de lavere indkomstgrupper ender i en lavere indkomstgruppe end deres fædre. I den nederste indkomstgruppe er der umiddelbart ikke nogen tendens.

Til sidst undersøger vi transitionsmatricerne for den geografiske opdeling. Her opdeles datasættet igen efter størrelserne på byerne, som forældrene kommer fra. Resultaterne af denne undersøgelse ses i figur 2.5. Antallet af observationer ses under figuren.

Figur 2.5 – Transitionsmatricer for opdelinger efter størrelse af byer.



Note: Figuren viser sandsynligheden for at børnene ender i samme indkomstgruppe som deres forældre, når der opdeles efter forskellige størrelser af by i DLSY. Antallet af observationer: KBH=67, forstad KBH=104, Stor by=328, Lille by=242, Landsby=236 og Landdistrikt=176.

Undersøgelsen af transitionsmatricerne for opdelingen efter størrelse af byer giver blandede resultater. I nogle af opdelingerne ses der en tendens for lav mobilitet i bunden og toppen af indkomstgrupperne, mens det er mere eller mindre tilfældigt i andre opdelinger. Som udgangspunkt kan det være svært at

forklare hvorfor, der er forskellige tendenser i de forskellige opdelinger. I forhold til transitionsmatricen med det samlede datasæt, ser vi generelt nogle mere tydelige tendenser i mobiliteten, når vi opdeler datasættet. Dette kan være med til at give en bedre forståelse for hvilke grupper, der har en høj eller lav mobilitet.

Overordnet set er det altså muligt at lave transitionsmatricer for det danske datasæt DLSY og dermed også muligt at lave generelt i små surveys. Som udgangspunkt er der ikke noget, der fortæller os hvor statistisk signifikante resultaterne er. Litteraturen viser dog, at transitionsmatricerne lige såvel som IGE og rank-rank korrelationer kan være påvirket af life cycle bias. Nybom & Stuhler (2016) viser, at hvis forældrenes indkomst bliver målt enten for tidligt eller for sent, kan det medføre, at mobiliteten overdrives i enderne af matricen. Det kan være svært at udlede noget generelt ud fra transitionsmatricerne, eftersom de kun siger noget om forholdet i det udvalgte datasæt. Der vil generelt være en større måleusikkerhed, hvis datasættet indeholder få observationer.

2.6. Sammenligning af DLSY og andre surveys

I dette afsnit vil vi sammenligne vores deskriptive undersøgelse af den intergenerationelle mobilitet i det danske survey DLSY i forhold til de præsenterede surveys og undersøgelser i afsnit 2.3. Formålet med dette er at få en forståelse for, om de metodiske problemer som vi har oplevet i analysen, er unikke til DLSY, eller om det er generelle problemer ved at undersøge den intergenerationelle mobilitet i surveys. Vi forsøger hermed også at besvare det første underspørgsmål, som lyder: Hvilke metodiske problematikker opstår der i forbindelse med at estimere den intergenerationelle mobilitet deskriptivt i observationsbegrænsede surveys (IGE, rank-rank og transitionsmatricer)?

Vi har i vores deskriptive analyse af den intergenerationelle mobilitet i DLSY oplevet flere metodiske problemer. Nogle af problemerne er almindelige, når det sammenlignes med andre lignende undersøgelser, mens andre er mere unikke i forhold til DLSY. Generelt er det muligt i litteraturen og vores analyse at løse den metodiske problematik i forhold til, at det er bedst at måle indkomsten, når individerne er i sluttyverne til omkring midten af fyrrerne. I vores analyse kan det være problematisk, at gennemsnitsalderen for børnenes indkomstdata er på omkring 32 år, mens forældrenes er 38 år. Derudover er børnenes indkomstdata også målt på forskellige tidspunkter af børnenes liv, hvilket medfører en vis usikkerhed, fordi individerne har forskellige vilkår.

Det viser sig at være et generelt problem i både litteraturen og i vores analyse, at der er få år med indkomstdata i surveys. I DLSY er der kun ét år med indkomstdata, hvilket også er tilfældet med GSS og BCS. Der findes dog også surveys, som indeholder flere års indkomstdata (NCDS, NLS og PSID),

men disse er stadig langt fra at indeholde nok år med indkomstdata i forhold til, hvad der anbefales i litteraturen for at minimere bias mest muligt (Mazumder, 2005). Derudover findes der også nogle problematikker i forhold til, at nogle surveys bruger tvivlsomme metoder til at måle indkomsten. Der vil generelt altid være en del tvivl omkring validiteten af målene for indkomst, eftersom folk ikke altid vil svare helt rigtigt, når de bliver spurgt om deres indkomst. I DLSY er indkomsten for forældrene lønindkomst og for børnene er det bruttoindkomst. Dette er et stort problem i forhold til at kunne undersøge den intergenerationelle mobilitet, eftersom de to indkomster ikke kan sammenlignes. Det viser sig dog at være et unikt problem til DLSY og finder ikke sted i de andre surveys. Generelt er det derfor ikke et problem, at indkomsten er målt på forskellige måder. Derimod er der i nogle surveys brugt tvivlsomme metoder til at måle indkomsten. I GSS og BCS er børnene blevet spurgt om deres families indkomst, hvilket betyder at målet er relativt usikkert. Et andet problem som opstår i forbindelse, med at undersøge den intergenerationelle mobilitet i surveys er, at der generelt er få observationer i forhold til undersøgelser udført med registerdata. Dette gør sig også gældende i vores analyse med DLSY og de andre undersøgelser vi har præsenteret fra litteraturen. Sammenlignet med de andre surveys er DLSY i den lave ende i forhold til observationer.

Der findes altså en del forskelle og ligheder mellem DLSY og de andre surveys, som vi har præsenteret. Generelt er der en del problematikker i forhold til at undersøge den intergenerationelle mobilitet i surveys, men de fleste undersøgelser fra litteraturen formår at finde statistisk signifikante resultater med relativt små standardfejl. Fælles for resultaterne er dog, at mobiliteten kan være overestimeret, eftersom de ofte ikke har mere end tre års indkomstdata tilgængeligt. I vores analyse af DLSY kan det som udgangspunkt være svært at sige noget om mobiliteten, eftersom at indkomsten for forældrene og børnene er opgjort forskelligt. Derudover finder vi også relativt store standardfejl, hvilket gør det svært at sige noget ud fra estimerne. Det vil altså sige, at der findes nogle problematikker, som er unikke til DLSY, hvilket gør at det er markant sværere at undersøge den intergenerationelle mobilitet i dette survey i forhold til andre. Generelt er det som udgangspunkt muligt at undersøge mobiliteten i andre surveys, dog med risici for at estimerne er påvirket af bias.

3. Kausale sammenhænge

Intergenerationel indkomstmobilitet er et felt, som rent forskningsmæssigt har en del ikke-udforskede områder. Om end det er vokset i popularitet gennem de seneste årtier, tager forskningen stadig små skridt mod at afklare de faktorer som spiller ind i forhold til intergenerationel mobilitet (Black & Devereux, 2011). I mange år har forskningen ved mere deskriptive studier fokuseret på at undersøge

korrelationerne mellem forældre og børn i forhold til indkomst og uddannelse, mens det i de senere år har fokuseret på at belyse de bagvedlæggende kausaleffekter. For nogle af de tidligste studier på området, forsøgte man at klarlægge, hvor stor betydning genetik havde i forhold til miljø. I dag er der blandt forskere på området bred konsensus om, at mindst en andel af den intergenerationelle korrelation vi kan observere, skyldes den genetiske arv, men hvor meget der udgøres af genetik og hvor meget der udgøres af miljømæssig opvækst, er stadig til debat (Black & Devereux, 2011).

Området har også nogle praktiske udfordringer i forhold til data, eftersom det er et emne, som sætter utroligt store krav til den data hvorpå undersøgelser baseres. Det kræver som regel en omfattende mængde og detaljeret data på individniveau. Problemet med små datasæt kan måske virke banalt i en dansk kontekst, da professionelle forskere som regel kan få adgang til de store administrative registre. Dette er imidlertid ikke lige så nemt for forskere i andre lande. Det er ikke en selvfølge at alle lande indsamler så store mængder data, som det er tilfældet i Danmark. Dette gør det naturligvis relevant at belyse de muligheder og begrænsninger et mindre survey-baseret datasæt kan give.

Målet med dette afsnit er at belyse de udfordringer, der kan opstå ved undersøgelser af kausaleffekter for intergenerationel mobilitet ved brug af små surveys. Til at påpege udfordringerne tager vi udgangspunkt i et eksempel, som har til formål at undersøge den kausale sammenhæng mellem uddannelse og intergenerationel mobilitet. Vi har valgt at undersøge kausaliteten mellem uddannelse og den intergenerationelle mobilitet fordi, at denne sammenhæng er blevet udforsket af flere i litteraturen (Pekkarinen et al 2009; Holmlund 2008; Meghir & Palme 2005), om end der ikke er opnået konsensus om størrelsen af de potentielle effekter. Vi har i vores analyse valgt at følge Pekkarinen et al. (2009), som undersøger effekten af en skolereform i Finland ved hjælp af en difference-in-differences model. I stedet for en skolereform undersøger vi Bistandsloven 1976, som havde til formål at øge den tidlige uddannelse hos børn.

Afsnittet starter med en gennemgang af den eksisterende litteratur indenfor området for at få en fornemmelse af, hvilke undersøgelser der tidligere er lavet i litteraturen af de kausale sammenhænge mellem uddannelse og den intergenerationelle mobilitet. Herefter beskrives de teoretiske og empiriske sammenhænge mellem tidlig uddannelse og intergenerationel mobilitet. Dette gøres for at forstå hvorfor, at en ændring i tidlig uddannelse hos børn, kan være med til at påvirke den intergenerationelle mobilitet. Dernæst redegøres der for Bistandsloven 1976, som er den udvalgte reform til at undersøge de kausale sammenhænge mellem tidlig uddannelse og den intergenerationelle mobilitet. Derefter beskriver vi den anvendte metode til at undersøge sammenhængen, som er en difference-in-

differences model. Afsnittet sluttet af med at kritisere en estimation på baggrund af små longitudinale surveys. Dette gøres ved at sammenholde de forskellige datakrav ved DiD og vores case med Bistandsloven 1976 og det danske survey DLSY.

3.1. Hvad siger litteraturen?

På tværs af litteraturen anerkendes det, at de næste skridt mod at afdække den intergenerationelle mobilitet, bør ligge ved en klarlægning af de kausale sammenhænge. En sådan klarlægning vil give indblik i, de muligheder magthavere har for at påvirke den intergenerationelle mobilitet og helt konkret forstå hvilken indflydelse politiske ændringer har på indkomstmobiliteten.

Ved en gennemgang af den eksisterende litteratur bliver det tydeligt hvordan omfanget af studier, der undersøger de kausale sammenhænge, er en del mindre end de mere deskriptive studier. Dette projekt bidrager derfor til den eksisterende litteratur ved at kombinere to grene af litteraturen, hvoraf den følgende gennemgang vil have fokus på moderne kausalanalyser.

Blandt studier der anvender DiD, har Pekkarinen et al. (2009) søgt at estimere effekterne af uddannelse på den intergenerationelle indkomstmobilitet. De ser derfor på en omfattende finsk skolereform fra 1972, der gradvist rulles ud i det finske skolesystem over de efterfølgende seks år. Den gradvise udrulning giver mulighed for opsætning af et kvasi-eksperiment, hvilket udnyttes ved brug af en difference-in-differences-model. Pekkarinen et al. (2009) viser at en signifikant reduktion i effekten af fædres indkomst på sønners indkomst på syv procentpoint, hvilket svarer til et fald på 23 pct. i den intergenerationelle indkomstelasticitet. Resultaterne peger derfor på en forbedring af skolesystemet for sekundær uddannelse som en metode til at øge den intergenerationelle indkomstmobilitet. Forfatterne påpeger afslutningsvist det fokus der især bør ligge på tidlig uddannelse som vigtig faktor i forhold til intergenerationel indkomstmobilitet og behovet for yderlig forskning på området.

Et senere studie af Havnes & Mogsted (2012) undersøger netop tidlig uddannelse i norske institutioner. Her ses der på en udvidelse i adgangen til nationale/kommunale børnehaver. Udvidelsen inkluderede både en forbedring i uddannelsesmateriale, gruppe størrelser, personale uddannelse og det fysiske miljø. Forfatterne udnytter etableringen af reformen i 1975, ved at se på de kommuner, hvor ændringen i de statslige tilskud til kommunale børnehaver er størst. De ser på de kommuner, hvor ændringen i børnehavernes kvalitet stiger mest (reformgruppe) og sammenligner med kommuner hvor ændringen i statslige tilskud er mindst (kontrolgruppe). Resultaterne viser et fald i den intergenerationelle indkomstelasticitet på 2.5 procentpoint. Derudover finder de, at lavindkomst familier

opnår den største gevinst ved udrulning af et offentligt nationalt børnehavesystem (læs: tidlig uddannelse). Fælles for både Pekkarinen et al. (2009) og Havnes & Mogstad (2012) er anvendelsen af nationale registerdata i hhv. Finland og Norge.

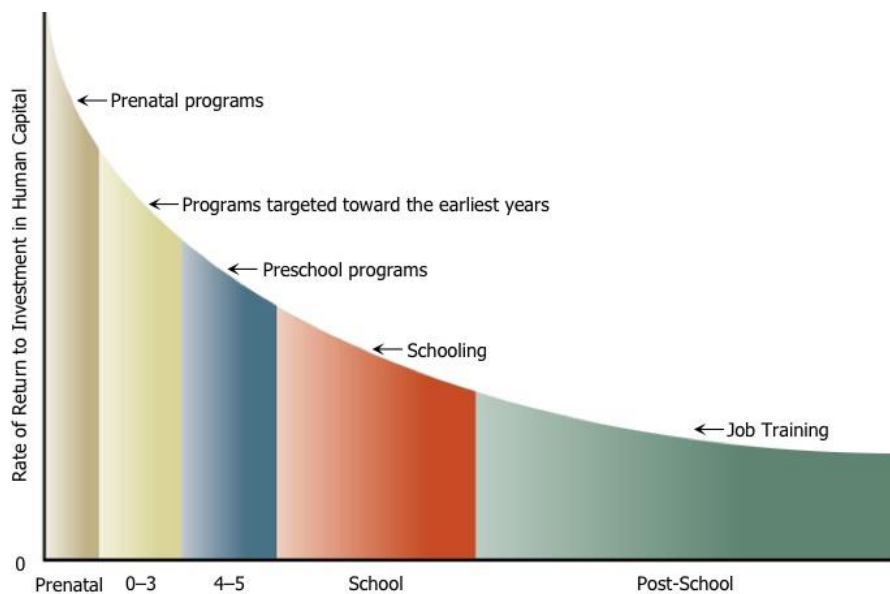
3.2. Tidlig uddannelse og intergenerational mobilitet

Dette afsnit har til formål at forklare, hvorfor en ændring i tidlig uddannelse hos børn kan påvirke den intergenerational mobilitet. Vi tager udgangspunkt i Heckmans forklaring af effekterne ved investering i human kapital. Dernæst redegøres der for forskellige udgivelser i litteraturen, som påviser, at tidlig uddannelse af børn har en stor indflydelse på deres fremtidige udfald heriblandt indkomst.

En af de centrale forskere indenfor området af effekterne ved tidlig investering i børns human kapital er Heckman, som bl.a. er udvikleren af den såkaldte *Heckman curve*. En figur der viser sammenhængen mellem netop afkastet ved investering i human kapital i forhold til alder. Hel centralt for Heckmans analyse er antagelsen om, at barndommen har flere stadier, og at menneskers evne til at lære afhænger af den viden og de evner der blev akkumuleret i forrige periode. Heckman beskriver altså børns uddannelse som en dynamisk proces, hvor kognitive og ikke-kognitive evner er en vigtig determinant, i forhold til de outcomes vi kan se hos børn senere i livet. Heckman ligger også vægt på, at de evner og den viden mennesker kan akkumulere i løbet af et liv ikke er identiske. Hver stadie har sine egne kendetegn, i forhold til de specifikke evner man lærer.

Figur 3.1 illustrerer den teoretiske effekt af investering i human kapital over en persons life cycle. Ved at plotte afkastet for investering i human kapital mod de forskellige stadier i en persons uddannelsesforløb får vi en aftagende produktivitets kurve. Første-aksen repræsenterer en persons alder, mens y-aksen repræsenterer afkastet ved investering i humankapital under den forudsætning, at der investeres den samme mængde i hvert stadie. Alt andet lige vil afkastet ved at investere f.eks. 100 kr. i børnehave-stadiet være højere end ved at investere 100 kr. i stadiet ved gymnasie eller erhvervsuddannelse (Cunha et al. 2006).

Figur 3.1 – Heckman-kurven



Kilde: Heckman, 2019 – Figuren viser forskellige afkast ved at investere i individers human kapital over deres life cycle.

Hovedpointerne i Heckmans forskning (Elango et al. 2015) er, at *Evner og intelligens er multipel* i den forstand, at kognitive evner (målt ved f.eks. IQ) ikke er et tilstrækkeligt mål for en persons intelligens. De ikke-kognitive evner f.eks. den interpersonelle, intrapersonelle eller eksistentielle intelligens er ligeledes vigtige faktorer i menneskers søgen efter et succesfuldt liv. Dette er alle evner, som tidlig uddannelse er med til at fremme.

Evner/intelligens er selvforstærkende og komplimenterende for hinanden. Imellem to tidspunkter i et hvert barns liv, t og $t + 1$, vil beholdningen for et barns evner altid bygge på den tidligere beholdning, med andre ord er den selvforstærkende. Derudover er den viden et barn modtager i de forskellige stadier komplementerende, hvilket vil sige at det som et barn lærer i tiden t kan kombineres med den viden barnet modtager i $t + 1$, og har altså større værdi end den separate viden hver for sig.

Evner er komplimenterende for investering. Ved at investere i tidlig uddannelse vil det øge kapaciteten i forhold til fremtidig indlæring og dermed produktiviteten for fremtidig investering. Af den grund er det f.eks. mere optimalt at investere i børn med evner, der ligger under gennemsnittet i en tidlig alder, frem for at indhente de manglende evner senere i livet (Elango et al. 2015).

Litteraturen indeholder også en del forskning indenfor området, som påviser, at tidlig uddannelse har stærke langsigtede effekter på individers fremtidige liv. Case & Paxson (2006) og Currie & Thomas (2001) viser, at børnehavebørn med lav kognitiv udvikling generelt klarer sig dårligere i skolen og

får en lavere indkomst som voksen. Et studie af Chetty et al. (2011) viser, at der er en høj korrelation mellem testscore fra børnehaven og indkomst som 27-årig, om man går på college, om man ejer et hus og pensionsopsparing. Heckman, Pinto og Savelyev (2013) viser, at det er mere effektivt at lave ændringer til børns uddannelse før de starter i skole i forhold til efter, de er startet i skole (hvilket også er beskrevet ovenfor). Yderligere beskriver de, at det kun er programmer, der starter før børnene, er blevet tre år gamle, som er med til at øge børnenes kognitive egenskaber på lang sigt.

Der findes altså en del forskning, som påviser, at en øgning i tidlig uddannelse hos børn, vil være med til at påvirke deres fremtidige liv positivt. I forhold til vores undersøgelse af de kausale sammenhænge mellem tidlig uddannelse og den intergenerationelle mobilitet er dette relevant fordi, vi undersøger effekterne af Bistandsloven 1976, som havde til formål at øge tidlig uddannelse hos børn i dagpleje. Hvordan reformen skulle øge den tidlige uddannelse, beskrives i det efterfølgende afsnit.

3.3. Bistandsloven 1976 – En reform for tidlig uddannelse

I dette afsnit beskriver vi, hvordan Bistandsloven 1976 er udformet og hvorfor denne reform er med til at øge den tidlige uddannelse hos børn. Formålet er at vise, hvordan reformen kan bruges som case til at påvise svagheder i studier af kausale sammenhænge mellem tidlig uddannelse og intergenerationel mobilitet i longitudinale surveys.

For at undersøge kausaliteten mellem intergenerationel mobilitet og uddannelse, har vi valgt at fokusere på lovændringer målrettet det vi kalder *tidlig uddannelse* og som det danske Børne- og Socialministerie definerer som institutioner til børn før folkeskolealderen (Socialministeriet, 2019). For de danske institutioner gælder det vuggestue, dagpleje og børnehave. Baggrunden for at vælge netop tidlig uddannelse har sit fundament ved Heckman-kurven, den store mængde forskning, som viser, at der er store effekter ved at øge den tidlige uddannelse hos børn samt den grundlæggende teori beskrevet i Becker og Tomes (1979), (1986) og Solon (2004). Vigtigheden af tidlig uddannelse fremgår derfor tydeligt ud fra litteraturen og teorier, både fra et socialt retfærdighedsprincip, hvor alle børn i et samfund får en mere lige start på livet, men også fra et effektivt økonomisk synspunkt.

Bistandsloven blev vedtaget i 1974 med politisk flertal fra Socialdemokratiet og Venstre og trådte i kraft d. 1. april 1976. Formålet med loven var at samle love omkring social forsorg og støtte. Baggrunden for at ændre lovgivningen var en forskydning af borgernes opfattelse af velfærdsstatens sikkerhedsnet. Statens ydelser skulle ikke længere gives som retsbaseret hjælp, men i stedet som behovsstyret hjælp. Hensigten var derfor at velfærden i højere grad kom alle borgere til gode, og at der

ikke længere blev skelnet mellem selvforskyldt og uforskyldt nød i forbindelse med de sociale ydelsers tildeling (Farbøl et al. 2018).

Bistandsloven var en hovedhjørnesteen i 1970'ernes socialreform, og lignende lovændringer så man også i Sverige i 1950'erne og i 1960'erne i Norge. Disse typer reformer er sidenhen anset som en fundamental del i opbyggelsen af det moderne velfærdssamfund, og med præcis det formål at give alle borgere lige muligheder (Pekkarinen et al. 2009).

Et led i bistandsreformen var etableringen af den kommunale dagpleje og herunder normeringen pr. plejer i institutionen³. Reformen bestemmer at, der for kommunale dagplejer maks. må være fem børn pr. plejer, hvor det førhen var syv børn (FOA, 2019). Normeringen for daginstitutioner i forhold til børns trivsel, evner og kompetencer er et stort empirisk område, der ligger uden for dette projekts ramme. Danmarks Evalueringsinstitut (EVA, 2017) opsummerer følgende:

- Normering og gruppestørrelse kan have væsentlig betydning for barn-voksen-interaktion.
- Normering i dagtilbud kan have betydning for børnenes sociale og kognitive kompetencer.
- Gruppestørrelsen har betydning for interaktioner og læring.
- Gruppeorganiseringen har betydning for børns leg og læringspotentialer.

Med udgangspunkt i den eksisterende forskning kan vi antage, at normeringen har betydning for børns muligheder for indlæring. Bistandsloven kan derfor ses som et positivt skridt mod en bedre uddannet befolkning. Vi vil derfor potentielt set kunne måle effekten af tidlig uddannelse og dermed undersøge de kausale sammenhænge mellem tidlig uddannelse og den intergenerationelle mobilitet ud fra denne reform.

Når man undersøger de kausale sammenhænge mellem uddannelse (eller andre variable) og den intergenerationelle mobilitet i surveys er en af problematikkerne, at det kan være svært at finde en reform, som opfylder datakravene for at kunne lave undersøgelsen. Denne problematik har vi oplevet i vores udvælgelse af en reform og i vores analyse af reformen. Hvordan denne problematik og andre vedrørende reformudvælgelse har påvirket analysen forklares i afsnittet om datakrav (3.5).

3.4. Metode

I dette afsnit vil vi forklare, hvordan man metodisk kan undersøge effekterne af f.eks. en lovændring eller reform og dermed finde de kausale sammenhænge mellem to variable. Dette kan gøres med en

³ Normering angiver antallet af plejere pr. barn. En høj normering vil derfor betyde et højt antal plejere pr. barn.

difference-in-differences model, hvis metode forklares i dette afsnit. Den metodiske gennemgang gør det muligt at diskutere udfordringer i brugen af data fra surveys.

3.4.1. Difference-in-differences

Difference-in-differences (DiD) er en ofte anvendt metode til at estimere kausaleffekter inden for samfundsvidenskaberne. Helt simpelt kan metoden, som navnet antyder, bruges til at udregne estimater for de forskelle man ser på tværs af grupper/individer over tid. Metoden gør sig særdeles brugbar i forbindelse med bestemte typer lovændringer, hvor der er mulighed for at måle de konkrete effekter af en specifik reform. Til forskel fra deskriptive undersøgelser søger vi ved denne metode ikke kun at finde sammenhængen, men også kausaliteten mellem den afhængige og forklarende variabel. Simplificeret, kan vi med denne metode placere en retning for de sammenhænge man ellers ser i f.eks. multiple regressionsanalyser. Sagt med andre ord - er det $x \rightarrow y$ eller $y \rightarrow x$. Begreberne lovændring og reform bruges ofte i nedenstående afsnit og kan erstatte hinanden. Begreberne bruges om det fænomen, at populationen udsættes for en ændring, som er eksogen i forhold til et relevant outcome for populationen. Den primære kilde i dette afsnit er Angrist & Pischke (2008).

Det mest enkle setup for DiD metoden, vil være en situationen, hvor man observerer udfald for to grupper i to forskellige perioder. Den ene gruppe vil opleve at blive udsat for en ændring i anden periode, men ikke i første periode. Denne gruppe benævnes reformgruppen. Den anden gruppe udgør kontrolgruppen og udsættes ikke for en ændring i hverken første eller anden periode. Kontrolgruppen udgør med andre ord en situation, hvor lovændringen er fraværende.

Vi er altså interesseret i at kende effekterne af reformen D_i på den afhængige variabel Y_i . Vi får således ligning 3.1.

$$Y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \varepsilon_{it} \quad (3.1)$$

α_i opfanger de gruppe- eller individspecifikke effekter, dvs. de faste karakteristika for individet eller gruppen og som ikke ændres over tid. λ_t opfanger de tids-specifikke effekter.

I modellen eksisterer der to grupper $s = A, B$ og to tidsperioder $t = 0, 1$. Ændringen i den forventede værdi for Y i reformgruppen udtrykkes ved ligning 3.2.

$$E(Y|s = A, t = 1) - E(Y|s = A, t = 0) = \lambda_2 + \beta - \lambda_1 \quad (3.2)$$

For kontrolgruppen udtrykkes ændringen i den forventede værdi som ligning 3.3.

$$E(Y|s = B, t = 1) - E(Y|s = B, t = 0) = \lambda_2 - \lambda_1 \quad (3.3)$$

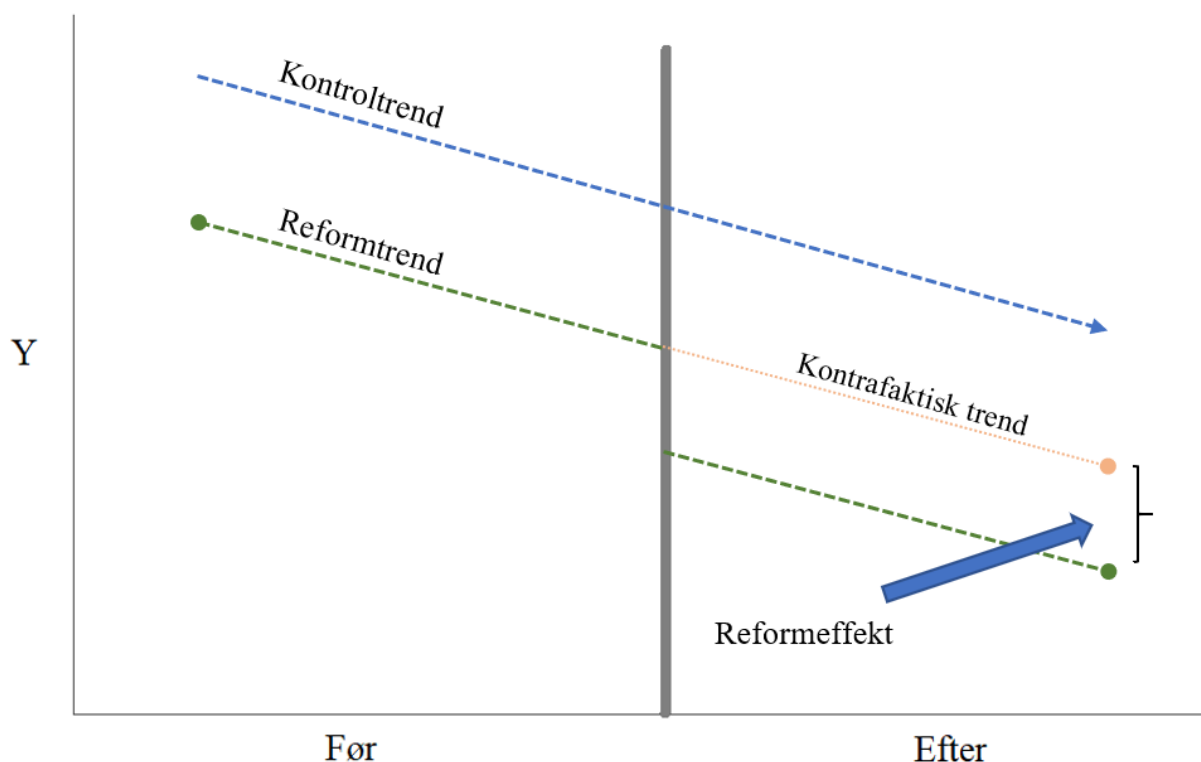
Herfra kan vi skabe DiD estimatoren β , som ses i ligning 3.4.

$$[E(Y|s = A, t = 1) - E(Y|s = A, t = 0)] - [E(Y|s = B, t = 1) - E(Y|s = B, t = 0)] = \beta \quad (3.4)$$

β måler derved effekterne af reformen, som beregnes ved at tage den gennemsnitlige ændring i reformgruppen minus den gennemsnitlige ændring for kontrolgruppen. Dette fjerner derved et potentielt bias mellem de to grupper i anden periode, som skyldes de permanente forskelle der er mellem reform- og kontrolgruppe. Det fjerner derudover også det bias, som kan eksistere ved målinger over tid, som påvirker gruppernes outcome ens.

Figur 3.2 illustrerer metodens grundantagelser. Figuren viser værdien af den afhængige variabel over tid for hver gruppe. Den vertikale linje indikerer at reformen træder i kraft. Den første antagelse i metoden er, at reformen ikke ændrer trendlinjen for kontrolgruppen. Reform-gruppen oplever herimod en forskydning af trendlinjen efter indførelse af reformen. Det er en essentiel del af DiD metoden, at vi kan antage parallelle trends for de to grupper i fraværet af reformen. Valideringen af modellen kan foregå ved at undersøge, hvorvidt der ses parallelle trends mellem grupperne forud for reformen (Angrist & Pischke, 2008; Mailman School of Public Health, 2019).

Figur 3.2 – Grafisk illustration af difference-in-differences



Note: Figuren illustrerer trendlinjen for kontrol- og reformgruppen og viser effekten ved indførelsen af en reform.

Identifikationsantagelser for den ikke-parametriske standardmodel.

En forudsætning for at kunne estimere kausaleffekterne af f.eks. en lovændring er, at en række antagelser opfyldes for modellen. Vi vil nu se nærmere på disse antagelser.

Den første og mest afgørende antagelse er 'common trends' (CT) også kendt som parallelle trends. Denne antagelse beskriver hvordan, at de gennemsnitlige ændringer for reform- og kontrolgruppen skal være ens. Dette er derfor afgørende for DiD-metoden og angiver hvordan, at trendlinjerne for både reform- og kontrolgruppen ville følge den samme parallelle udvikling i fravær af en lovændring. Det sætter samtidig krav om, at alle relevante kontrolvariable inkluderes i modellen. En situation hvor f.eks. region eller uddannelsesmæssig baggrund påvirker trendlinjerne til at bevæge sig modsat hinanden, vil være et brud på CT-antagelsen. Derudover giver det også nogle klare udfordringer i forhold til udvælgelsen af kontrolgruppen. Det er typisk dette som gør det problematisk at skabe et korrekt og unbiased DiD-setup. Vi vender tilbage til denne udfordring når casen i dette projekt gennemgås.

Den næste antagelse er 'stable unit treatment value assumption' (SUTVA). Denne antagelse består af to dele hvoraf den første bestemmer, at der kun kan være ét potentielt outcome for alle individer i populationen. Alle individer kan derved kun observeres som enten kontrolgruppe eller reformgruppe. Anden del bestemmer, at den ene gruppe ikke må påvirkes af den anden, eftersom potentielle outcomes kun kan observeres for enten kontrolgruppe eller reformgruppe. Brud på denne antagelse vil betyde, at vi ikke længere kan isolere effekten af reformen men, at andre ændringer har indflydelse på outcome. Et eksempel der giver brud på antagelsen, ville være en folkeskole, hvor der indføres lektiehjælp for enkelte elever i forsøg på at øge deres karakterer(reformgruppe). De elever som modtager lektiehjælp vil i løbet af dagen interagere med de elever som ikke modtager lektiehjælp(kontrolgruppe) og måske videregive den viden de har fået gennem lektiehjælpen. Fænomenet er kendt som spillovereffekter, og vil i dette tilfælde umuliggøre at identificere den "rene" effekt af lektiehjælp (læs: reformen). Løsningen på dette kunne være at gå fra elev-niveau til skole-niveau for at mindske den sociale interaktion mellem kontrol- og reformgruppe (Bloome, 2009).

Den næste antagelse omhandler eksogenitet (EXOG) for kontrolvariable i modellen. Antagelsen er ligesom SUTVA en standard antagelse for mikroøkonometriske kausalstudier. Her kan typisk opstå brud på antagelsen, når der måles på udfald i perioden efter en lovændring. Et eksempel kunne være en gruppe arbejdsløse, der udsættes for jobaktivering(reformgruppe). Her kunne man forestille sig, at en kontrolvariabel som f.eks. arbejdstilfredshed vil være endogent bestemt i det omfang, at den afhænger af, hvorvidt de arbejdsløse der udsættes for jobaktivering, får et arbejde. At bruge målet for

arbejdstilfredshed i perioden før jobtræningen giver ligeledes heller ikke en garanti for eksogenitet, eftersom de arbejdsløse med udsigt til jobaktivering kan ændre adfærd.

Den tredje antagelse 'no effect on the pre treatment' (NEPT) er mere konkret for DiD-metoden. Den minder om EXOG i den forstand, at den behandler de adfærdsmæssige ændringer, der kan opstå ved udsigten til en lovændring.

Forskellen ligger ved, at den i stedet omhandler eksogenitet for den afhængige variable. Et brud på denne antagelse ville f.eks. være situationen, hvor vi måler på outcome for arbejdsløshed, og hvor de arbejdsløse i reformgruppen stopper deres jobsøgning, da de har udsigt til jobaktivering og derved større sandsynlighed for at få et arbejde den vej igennem (Lechner, 2011).

Et kendt og ofte citeret eksempel er Card & Krueger (1993), som ved hjælp af DiD søgte at påvise kausaleffekterne ved en lovændring af minimumslønninger på statsniveau i USA. Lovændringen var derfor kun gældende i udvalgte stater og altså ikke på føderalt niveau. Studiets popularitet skyldes formentlig det faktum, at grundlæggende økonomisk teori inden for arbejdsmarkedsøkonomi dikterer, at i et konkurrerende marked vil en stigning i minimumslønninger flytte virksomheder længere op på den nedadgående efterspørgselskurve. Det vil derved øge arbejdsløsheden for netop de arbejdere man som udgangspunkt forsøger at hjælpe. Resultaterne viste, at dette ikke var tilfældet, men at arbejdsløsheden faldt i de stater hvor minimumslønninger steg. Undersøgelsen blev derfor hyppigt inddraget af de organisationer, der ønskede at øge minimumslønningerne. Omvendt blev studiet i høj grad kritiseret hos de organisationer, der var imod stigninger i minimumslønninger.

I setup'et hos Card & Krueger vil DiD metoden grundlæggende tage højde for, at ikke alle observerede ændringer skyldes den konkrete lovændring. I undersøgelsen ses der på to forskellige stater over to perioder⁴. Med DiD får vi mulighed for adskille de effekter, der kan relatere sig til den enkelte stat fra den aktuelle lovændring i undersøgelsen. Det kunne f.eks. være andre reformer på arbejdsmarkedet, som kan påvirke beskæftigelsen. Derudover adskiller metoden også de tidsrelaterede effekter fra lovændringen vi er interesseret i. Eftersom der i setup'et hos Card & Krueger (1993) ses på to forskellige tidspunkter, er det også forventeligt at der sker ændringer i samfundet over tid. Typisk vil det makroøkonomiske klima ændre sig og derfor have en indflydelse på beskæftigelsen i alle stater i USA. I ligning 3.1 udtrykkes disse gruppe- og tidsvarierende effekter ved hhv. α_i og λ_t .

⁴ De to stater er New Jersey og Pennsylvania, som observeres i hhv. februar 1992 og november 1992.

3.5. Datakrav for DiD i forhold til DLSY

Ved kvantitativ analyse af kausaleffekter og herunder ændringer over tid, er det også et krav, at den data hvorpå der analyseres indeholder et tidselement. Vi vil i dette afsnit berøre de forskellige typer af data, som kan anvendes ved DiD-metoden hovedsagelig med fokus på det data, som benyttes i dette projekt og hvilke problematikker det medfører at bruge data fra et survey. Derudover vil vi også beskrive de specifikke problematikker i forhold til brug af DiD med det danske survey DLSY. Afsnittet har til hensigt at beskrive de forskellige fordele og ulemper, for den vej igennem at muliggøre en diskussion omkring problematikkerne ved at anvende små surveys til at undersøge de kausale sammenhænge mellem uddannelse og den intergenerationelle mobilitet.

Der eksisterer tre typer data, som alle har observationer der ligger på forskellige tidspunkter:

- Repeated cross-sectional survey data
- Retrospective survey data
- Longitudinal survey data

Repeated cross-sectional survey data er, som navnet antyder, blot cross sectional data der observeres på forskellige tidspunkter. Informationer i denne type data kommer derfor ikke fra det samme sample/individ, men kan være aggregeret for f.eks. en kommune. Dvs. at der ved en årlig spørgeundersøgelse i f.eks. Aalborg kommune, vil være et nyt sample hvert år. Denne type data har mange begrænsninger i forhold til panel data, men opfylder kravene og vil typisk blive anvendt med aggregeret tal for en gruppe.

Retrospective survey data er data der indsamles ved at spørge individer om tidligere forhold, dvs. at hente informationer ud fra deltagernes hukommelse og altså ikke nuværende forhold. Denne type data er typisk en del af f.eks. en cross sectional survey. Informationerne kan være anvendelige, men kan også indeholde en del usikkerhed.

Longitudinal survey data også kendt som paneldata er observationer af individer på forskellige tidspunkter. Til forskel fra cross sectional data, er der her tale om de samme individer som følges over tid, eller rettere, i bølger. Personer som interviewes i bølge 1, er de samme som interviewes i bølge 2 osv. Paneldata er altså en kombination af cross sectional- og tidsseriedata. Paneldata har mange fordele, hovedsageligt fordi vi har mulighed for at følge de samme personer over tid. Så mens cross sectional data kan anvendes til at måle på de overordnede niveauforskelle, kan paneldata i stedet

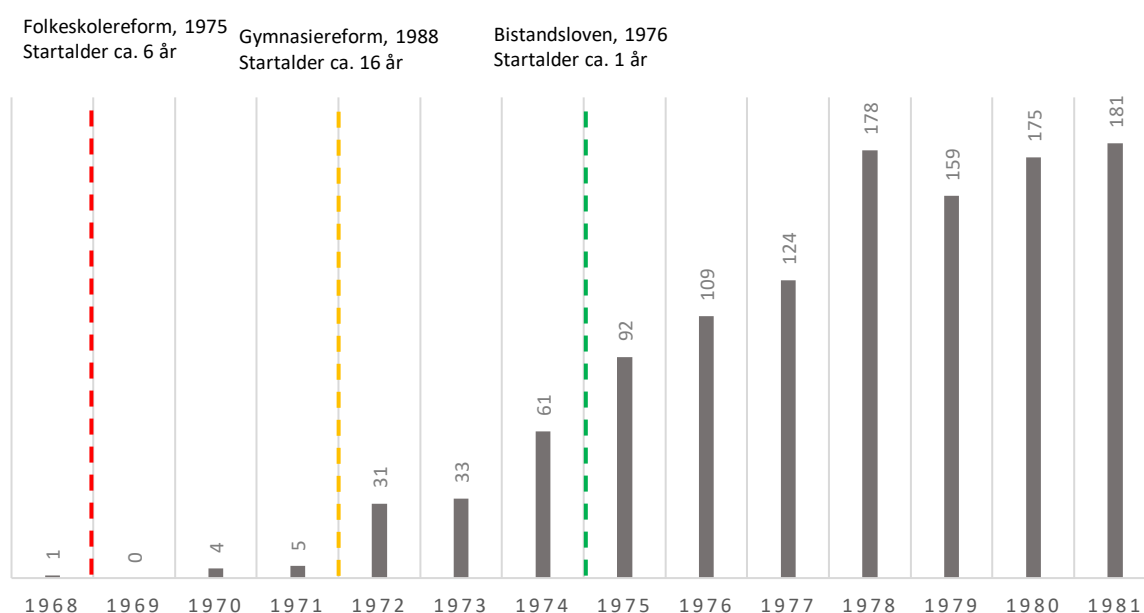
benyttes til at finde de underliggende ændringer. I relation til dette projekt, kan vi eksempelvis nævne effekterne af tidlig uddannelse i forhold til intergenerationel mobilitet.

Under *longitudinal survey data* findes også *cohort survey data* som anvendes i dette projekt. Her følger vi de samme personer fra et identisk tidspunkt i deltagernes liv. Som det nævnes i afsnit 2.4, følger vi i bølger interviews for en generation født i 1954 og efterfølgende deres børn. Datatypen gør sig utrolig brugbar når det kommer til at undersøge udviklingen for børn og generelt forskelle mellem generationer. Da kohorten er født på samme tid, vil man kunne adskille *cohort effects* og *age effects*. Hvor alderseffekter kan skyldes det faktum at f.eks. indkomsten for en person i løbet af livscyklus kan variere, så kan kohorteeffekter skyldes kulturelle, økonomiske eller teknologiske forskelle mellem generationer. Særligt alderseffekter er noget som kan have stor betydning i det datasæt som anvendes i dette projekt. Selvom forældregenerationen fra 1954 eliminerer alderseffekter, er det ikke tilfældet for den efterfølgende generation. Som det fremgår af figur 2.2, er børnene i undersøgelsen ikke født på samme tid. Det tidligst fødte barn i vores baseline model er fra 1968, mens det senest fødte barn er fra 1982. Børnene kan derfor økonomisk og socialt opleve helt forskellige tidspunkter i deres liv, da de bliver interviewet i 2010. Ligeledes kan kohorteeffekter også være aktuelle, f.eks. kan det økonomiske klima have skiftet og derved have indflydelse på indkomsten for børn født omkring 1968 kontra 1982 (UK Data Service, 2015).

Taler vi om de udfordringer et cohort survey kan give, har der fra starten af projektet eksisteret to markante problemstillinger. Den første udfordring er at finde en relevant lovændring med et klart formål, der er foreneligt med uddannelse som variabel, og som tidsmæssigt stemmer med kohortens alder. Den anden udfordring er, at man skal finde en reform, som har nok observationer på begge sider af reformen. Angrist & Pischke (2008) viser i deres analyse 'Parallel Worlds', at der skal eksistere observationer før og efter en given ændring. Dette er ikke altid muligt, når man arbejder med surveys eftersom, at det tilgængelige data er fastlåst til nogle bestemte år, hvor individerne er blevet interviewet. Et eksempel der påviser disse problematikker, er gymnasireformen fra 1988, som indeholdte en omstrukturering af gren-systemet, der skulle sætte fokus på at øge fagligheden i gymnasiet. Rent tidsmæssigt ligger reformen for tidligt i forhold til populationen der indgår i DLSY. Det betyder, at der kun eksisterer 10 observationer hvor et individ har gennemført gymnasiet før reformen. Denne problematik illustreres i figur 3.3. Derudover har reformens effekter været empirisk debatteret, hvilket vanskeliggør at definere reformen som et positivt eller negativt stød til uddannelse. Som udgangspunkt er det ikke vigtigt om stødet er positivt eller negativt, men hvis reformen ikke har en "klar" effekt, er det svært at sige noget ud fra undersøgelsen. En anden reform som derimod har en klar

effekt, er folkeskolereformen fra 1975. Denne reform var med til at udvide folkeskolen fra at slutte i 7. klasse til at slutte i 9. klasse. Reformen minder på mange måder om den reform, som Pekkarinen et al. (2009) undersøger i Finland. Det kunne derfor have været interessant at undersøge en lignende reform for at sammenligne resultaterne mellem de to. Som figur 3.3 viser, er dette dog ikke muligt eftersom, at der kun ville være én observation tilgængelig før reformen, når man bruger DLSY.

Figur 3.3 – Observationer i DLSY i forhold til udvalgte reformer.



Note: Figuren viser observationer i forhold til reformernes start. Kigger vi på folkeskolereformen indikerer den stiplede røde linje, at kun én person i vores population har gennemført folkeskole før/delvist før reformen. Resten ligger efter reformens indførelse.

En tredje udfordring ligger i udvælgelsen af kontrolgruppen. Det er en essentiel del af DiD at finde en repræsentativ kontrolgruppe, der opfylder antagelserne for metoden. Bistandsloven trådte i kraft over hele landet d. 1. april 1976. Det giver derfor en udfordring i forhold til at skabe en klar kontrolgruppe. For at kunne opstille en kontrolgruppe, skal der være en gruppe, som ikke bliver påvirket af reformen og derfor kan bruges som sammenligning. Det er en essentiel del af DiD metoden at kunne opstille en kontrolgruppe, som man kan sammenligne med for at finde den faktiske effekt af reformen. Det kan derfor være en udfordring med reformer eller lovændringer, som påvirker hele landet. En måde hvorpå Pekkarinen et al. (2009) undgik dette problem var, at reformen blev som udgangspunkt udført over hele landet, men det skete på forskellige tidspunkter. De kunne derfor sammenligne de forskellige grupper. Der findes også måder, hvorpå vi kan opstille forskellige alternative kontrolgrupper med Bistandsloven 1976. Det ville f.eks. potentielt set være muligt at opstille en kontrolgruppe,

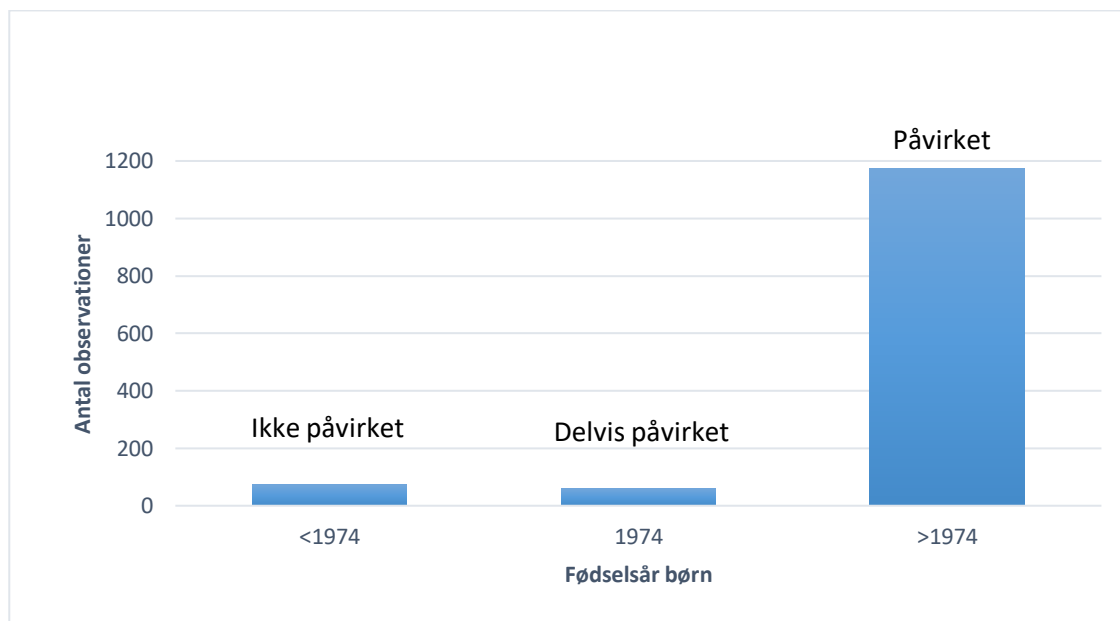
hvis man havde data over hvor mange kommuner, som af forskellige årsager allerede havde fem børn pr. dagplejer før reformen. På den måde ville man kunne lave en kontrolgruppe eftersom, at de i princippet ikke bliver påvirket af reformen.

En anden essentiel del af DiD metoden er antagelsen om parallelle trends, som antager, at den påvirkede gruppe og kontrolgruppen har den samme trend, hvilket også er forklaret i metodeafsnittet. Denne antagelse kan være svær at teste for, men kan testes ved at lave en ”visuel test”, som følger grupperne over tid og tjekker om de følger den samme trend over tid. Det stiller krav til en klar og stor kontrolgruppe samt længden vi følger gruppen. Vi kan med det danske survey DLSY ikke lave denne form for test, eftersom der ikke er mere end et år med indkomstdata.

Når man estimerer effekten af en reform med en DiD model, gøres det rent metodisk ved at lave et interaktionsled mellem en dummy, som definerer tidspunktet for reformen og en dummy som definerer gruppen, der bliver påvirket af reformen (opstiller kontrolgruppe). Dette kan være problematisk, hvis der er en begrænset mængde observationer i det anvendte survey og hvis indførelsetidspunktet af reformen passer skævt med det tilgængelige data. Det danske survey DLSY gør det svært at opstille en kontrolgruppe eftersom, at reformen bliver indført på samme tid over hele landet men også fordi, at størstedelen af kohorten bliver påvirket af reformen. Derudover er der også en skæv fordeling af observationer på begge sider af reformen. Det antages, at børnene starter i dagpleje, når de er et år gamle og at de stopper, når de er tre år gamle⁵. Dette betyder, at hvis et barn er født i 1974, vil barnet nå at få et år i dagplejen uden reformen og et år i dagplejen med reformen. Hvis et barn er født i 1975, vil barnet gå i dagpleje med reformen og hvis et barn er født før 1974, vil barnet ikke opleve reformen. I figur 3.4 kan det ses, at langt størstedelen af børnene er født efter 1974 og derfor udelukkende går i en dagpleje, hvor reformen er indført. Det ses også, at der er markant færre observationer i gruppen, som ikke oplever reformen og i gruppen som oplever begge dele. Dette vil være problematisk i opstillingen af de to dummys eftersom, at der er markant færre individer, som ikke bliver påvirket af reformen i forhold til individerne, som bliver påvirket af reformen.

⁵ Baseret på 2018 tal fra Danmarks Statistiks registerdata er størstedelen af danske børn tilknyttet en daginstitution fra de er ca. 1 år. Derudover var der i 2017 en %-indskrivning af børn i børnehavealderen (3-5 år) på 88,6 pct.

Figur 3.4 – Observationer i DLSY i forhold til Bistandsloven 1976.



Note: Figuren viser hvor mange af observationer i DLSY, der bliver påvirket af Bistandsloven fra 1976. Vi antager, at børnene starter i dagpleje, når de bliver 1 år og har en gennemførelsestid på to år. Derfor er observationer før 1974 ikke påvirket, observationerne fra 1974 er delvis påvirket og observationerne efter 1974 er påvirket.

En problematik som er specifik i forhold til vores case, er antagelsen om, at alle børn der indgår i populationen, gennemfører et dagplejeforløb fra de er 1-3 år. Det stiller unægteligt spørgsmål til hvorvidt denne antagelse er korrekt. I 2014 var der for børn i alderen 0-2 år en indskrivningsprocent på ca. 20 pct. i dagpleje og for vuggestuebørn ca. 40 pct. Disse indskrivningsprocenter er formentlig langt højere end ved oprettelsen af den kommunale dagpleje i 1976. Det vil derfor være urimeligt at antage at alle børn i vores population gennemføre et dagplejeforløb. Alternativt ville det være mere korrekt at antage at alle børn i populationen gennemførte et børnehaveforløb, hvor indskrivningsprocenten i 2014 er på 84 pct. (Glavind & Pade, 2018). Det giver på sin vis en udfordring i forhold til målestøj, hvor observationer fra de individer der hverken før eller efter reformen gennemfører et dagplejeforløb. Udvælgelsen af en korrekt population, er derfor problematisk da en andel af personer i survey'en på intet tidspunkt er under risiko for at blive påvirket af reformen. Problemet er en markant begrænsning i forhold til DiD-metoden og umuliggør helt grundlæggende udarbejdelsen af en korrekt estimering for reformens effekter. Skulle denne problematik løses ville det kræve data fra f.eks. registeret *Indskrevne i daginstitutioner*, som bygger på IT-systemer for opkrævning af forældrebetalingen i daginstitutioner og dagpleje. Her ville der være muligt at se på ydelsesform for børn ved opgørelsestidspunktet og derved kunne censurerer for individer som er uden for risiko i forhold til reformen.

Det fremstår tydeligt, at der er en del problematikker i forhold til at bruge en DiD model med det danske survey DLSY. For at opsummere er det først og fremmest i princippet ikke muligt at bruge DiD metoden med DLSY eftersom, at datasættet ikke indeholder indkomstdata for mere end ét år og er altså derfor ikke paneldata, som metoden kræver. Dernæst er det en udfordring at opstille en kontrolgruppe, hvilket er problematisk, fordi det er en essentiel del af metoden der gør det muligt at måle effekten af reformen på den intergenerationelle mobilitet. Det kan muligvis lade sig gøre at opstille alternative kontrolgrupper, hvis man har informationer om, hvor mange børn der er i dagplejerne før reformen. Hvis nogle dagplejer allerede har fem børn pr. voksen, kan disse bruges som kontrolgrupper eftersom, at de ikke vil blive påvirket af reformen. En anden problematik som opstår i forbindelse med brug af DiD og surveys er, at det kan være svært at finde en reform, der passer i forhold til det tilgængelige data. Først skal der findes en reform, som har en klar effekt i forhold til hvad man undersøger og dernæst skal den passe i forhold til årene, hvor individerne i kohorten er blevet interviewet. Yderligere skal reformen passe således, at der er nogenlunde den samme mængde observationer af individer, som bliver påvirket og ikke påvirket. Til sidst skal reformen også være designet på en måde, så det er muligt at opstille en kontrolgruppe.

4. Diskussion

I det følgende afsnit vil vi sammenholde resultaterne fra analyserne med de opstillede spørgsmål i starten af projektet. Vi vil forklare meningen med resultaterne og sammenligne med andre lignende studier. Yderligere vil vi diskutere eventuelle begrænsninger i forhold til resultaterne og komme med forslag til fremtidig forskning indenfor området.

Vi har i projektet opstillet to underspørgsmål for at hjælpe os med at undersøge de metodiske problematikker i forhold til at undersøge den intergenerationelle mobilitet med data fra surveys. Det første underspørgsmål lyder: Hvilke metodiske problematikker opstår der i forbindelse med at estimere den intergenerationelle mobilitet deskriptivt i observationsbegrænsede surveys?

For at kunne svare på det første underspørgsmål har vi redegjort for de mest anvendte metoder i litteraturen til at estimere den intergenerationelle mobilitet (IGE, rank-rank korrelation og transitionsmatricer) med fokus på eventuelle problematikker ved at bruge disse. Her fandt vi, at de mest gængse problematikker relaterer sig til at kunne måle den permanente indkomst for generationen af forældre og børn. Persistent transitoriske shocks beskriver problematikken ved at estimere den intergenerationelle mobilitet med få års indkomstdata, mens life cycle bias beskriver vigtigheden ved at

måle indkomsten, når individernes indkomst er stabiliseret (som regel omkring 30-45 år). Hvis ikke det er muligt at imødekomme disse problematikker viser det sig, at estimerne højst sandsynligt vil være påvirket af bias, som betyder, at estimerne af mobiliteten underestimeres (Solon 1992; Zimmerman 1992; Mazumder 2005; Baker & Solon 2003; Haider & Solon 2006). Litteraturen har indtil for nyligt i høj grad fokuseret på at undersøge bias i estimatet for IGE, mens der har været mindre fokus rank-rank korrelationer og transitionsmatricer. Det er derfor umiddelbart mindre klart, hvilke problematikker disse mål indebærer. En ny rapport fra Nybom & Stuhler (2016) viser dog, at både rank-rank korrelationer og transitionsmatricer til en vis grad påvirkes af de samme problematikker som IGE.

Problematikkerne ved at bruge metoderne stiller nogle krav til det data, der bruges til at estimere den intergenerationelle mobilitet. Den generelle begrænsning af data indenfor området i mange lande har medført, at vi har undersøgt, om det danske survey DSLY opfylder kravene. Derudover præsenterer vi de mest anvendte surveys i litteraturen og undersøgelser som bruger disse. Dette giver os muligheden for at kunne sammenligne med vores resultater fra DSLY og dermed få en bredere konklusion. En generel problematik der er forbundet, med brug af surveys er, at der er usikkerhed omkring, hvorvidt individerne svarer præcist til spørgsmål omkring indkomst. En anden problematik som kan opstå i forbindelse med brug af surveys er, at spørgsmålene i de forskellige interviewrunder kan være forskelligt formuleret. I det danske survey DSLY finder vi, at indkomsten er opgjort på to forskellige måder for henholdsvis forældrene og børnene. Samme problematik opstår ikke i de mest anvendte surveys fra litteraturen og noget tyder derfor på, at det kan være en unik problematik i forhold til DSLY. Det er dog et godt eksempel på, hvilke udfordringer forskere kan støde på ved brug af surveys og lignende problematikker kan være tilstede i mindre surveys end de inkluderede surveys i dette projekt. I mange lande betyder den begrænsede surveydata, at forskere er nødsaget til at estimere individernes indkomst, fordi de tilgængelige surveys ikke indeholder indkomstdata. Dette gør Narayan et al. (2018) for at kunne måle den intergenerationelle mobilitet for 75 lande. De bruger en økonometrisk metode (TSTSLS), som estimerer individernes indkomst ud fra deres uddannelsesniveau. I teorien kunne dette være en god måde, hvorpå forskere kan komme udenom problemet med manglende indkomstdata. I praksis er metoden dog forbundet med stor usikkerhed, som betyder, at forskere som udgangspunkt ikke burde sætte deres lid til denne metode. En anden udfordring i forhold til brug af surveys er, at der kan være usikkerhed omkring nøjagtigheden af den målte indkomst. I de amerikanske surveys GSS og NLS er børnene blevet spurgt om deres forældres indkomst, hvilket medfører en stor usikkerhed. I vores analyse af den intergenerationelle mobilitet i DSLY finder vi, at

det er udfordrende at forhindre den bias, der relaterer sig til persistent transitoriske shocks eftersom, at der kun er et år med indkomstdata tilgængeligt. Den samme tendens er også tilstede i de mest anvendte surveys fra litteraturen, hvor de fleste undersøgelser bruger indkomstdata fra tre eller færre år (med undtagelse af PSID). Det tyder derfor på, at det er et generelt problem, som er svært at løse ved brug af surveys. Derimod finder vi, at det generelt er muligt at tage højde for problematikken ved life cycle bias eftersom, at indkomsten er tilgængeligt for forældre og børn omkring 30-45 års alderen. Det kan dog være problematisk, hvis en begrænsning af observationer gør, at det tilgængelige indkomstdata repræsenterer forskellige aldersgrupper. Dette er en begrænsning, som vi finder i DSLY. I relation til observationsbegrænsninger finder vi også, at et generelt problem ved brug af surveys er, at der er relativt få observationer tilgængeligt. Vi finder i DSLY, at opdelinger efter far og søn og far og datter samt geografi er problematisk fordi, at datasættet indeholder relativt få observationer. Den samme tendens finder også sted i de mest gængse surveys fra litteraturen, hvor flere surveys indeholder færre end 1000 observationer. Estimerne dækker derfor over en begrænset mængde af befolkningen.

Overordnet set finder vi, at det generelt er muligt at måle den intergenerationelle mobilitet deskriptivt, men at brug af surveys gør det relativt svært at korrigere for bias.

Det andet underspørgsmål lyder: Hvilke metodiske problematikker opstår der i forbindelse med at finde de kausale sammenhænge mellem den intergenerationelle mobilitet og tidlig uddannelse? For at kunne besvare spørgsmålet har vi opstillet en case, hvor vi ved hjælp af metoden difference-in-differences søger at påvise problematikkerne ved estimering af kausaleffekter. I stedet for at estimere de faktiske effekter diskuterer vi problematikkerne, der kan opstå med begrænset data i surveys. Ved DiD metoden er det muligt at påvise konkrete effekter af politiske reformer. Vi har valgt at tage udgangspunkt i Bistandsloven 1976, der bl.a. havde til hensigt at øge kvaliteten for tidlig uddannelse i Danmark. For at forstå de begrænsninger der kan opstå ved brug af DiD har vi gennemgået den metodiske fremgang og præsenteret den teoretiske model.

Herefter har vi gennemgået datakravene ved at bruge en DiD model generelt og sammenlignet med vores case Bistandsloven 1976. Overordnet finder vi, at det kan være problematisk at bruge surveys, når man undersøger kausaleffekter, fordi det er svært at finde en reform der er kompatibel med datasættet. Ser vi på DiD, har metoden en række hovedantagelser der skal opfyldes for at kausaleffekterne kan påvises. Det stiller høje krav til den data der benyttes i analysen og kan i praksis være svært at

opfyldte. Den mest essentielle antagelse er antagelsen om parallelle trends, som sætter krav til at reformgruppen og kontrolgruppen følger den samme trend frem mod en given lovændring. Herunder følger EXOG og NEPT der stiller krav til gruppernes adfærd frem mod lovændringen samt eksogenitet for kontrolvariable ved lovændringer. Derudover beskriver SUTVA hvordan, at metoden kræver en klar opdeling mellem outcome for individer i populationen og hvordan der ikke må eksistere interaktion mellem reform og kontrolgruppe.

I forhold til DLSY som vores case er baseret på, møder vi nogle klare begrænsninger. Det kommer til udtryk ved strukturen i surveyen, hvor der kun eksisterer observationer for et enkelt år hos børnene. Det giver en klar problematik i forhold til hovedantagelsen om parallelle trends, fordi der med observationer for kun et enkelt år ikke kan observeres en trend for de to grupper. Dette giver samtidig brud på EXOG og NEPT der også stiller krav til observationer over tid. Et generelt problem med surveys er, at det kan være svært at definere populationen. Økonometrisk giver det en udfordring i forhold til om populationen reelt er repræsentativ. Ved DLSY kan vi ikke afgøre hvorvidt populationen reelt udgøres af personer under risiko for at opleve reformen, hvilket skyldes manglen på informationer omkring individers tilknytning til danske dagplejer. Derfor må vi antage at pga. de lave indskrivningsprocenter i 1970'erne, vil en forholds lille del af vores population have benyttet sig af dagplejesystemet.

Som nævnt kan det være svært at finde en reform der tidsmæssigt stemmer med kohortens alder. Gymnasireformen 1988 og folkeskolereformen 1975 er begge eksempler på reformer der tidsmæssigt ligger skævt i forhold til populationens alder. DLSY umuliggør derfor at anvende de to reformer og et komparativt studie med Pekkarinen et al. (2009) kan derfor ligeledes heller ikke lade sig gøre, hvilket ville være oplagt, hvis man ser på de fællestræk der eksisterer mellem folkeskolereformen 1975 i Danmark og den finske skolereform i 1972.

Ud fra vores case kan vi ikke opstille korrekte kontrolgrupper, da reformen gør det problematisk at definere, hvem der i populationen er upåvirket af reformen og på den vis gør det svært at skabe individ og tids dummies. Da reformen indføres nationalt, opstår der ikke nogen "naturlig" kontrolgruppe, som kan selekteres. Alternativt vil man med information omkring normeringen i de danske kommuner, kunne udvælge de steder hvor normeringen er højest, og dermed have den mindste effekt. Dette vil give en kontrolgruppe, hvor reformen har ingen/begrænset effekt. Denne fremgangsmåde er bl.a. set i Havne & Mogstad (2012).

Ved at besvare de to underspørgsmål finder vi flere metodiske problematikker ved at undersøge den intergenerationelle mobilitet med surveys. Herudfra opstår spørgsmålet om hvorvidt, det er muligt at tage højde for de forskellige metodiske problematikker og hvordan dette gøres, når man undersøger den intergenerationelle mobilitet med surveys. Vores projekt giver ikke alle svarene på dette, men er med til at give et overblik over hvilke problematikker der generelt opstår, hvis forskere eller studerende skal lave en undersøgelse af den intergenerationelle mobilitet med et survey. Vi har i dette projekt undersøgt problemerne ved at bruge de mest anvendte mål for mobiliteten og udfordringer ved at finde kausalitet, hvorimod tidligere udgivelser fra litteraturen oftest fokuserer på ét mål af gangen. Der er efter vores kendskab ikke nogen udgivelser fra litteraturen, som berører problematikkerne ved at undersøge de kausale sammenhænge ved den intergenerationelle mobilitet med brug af surveys. Vi er derfor med til at bidrage til litteraturen ved at lave et projekt, som indeholder de forskellige problematikker samlet.

Resultaterne fra projektet er med til at give en forståelse for hvad og hvordan der kan siges noget om den intergenerationelle mobilitet, når man bruger surveys. Det er umiddelbart svært at løse problematikken med, at der er få års indkomstdata tilgængeligt, men hvis vi har information om, hvor meget bias der er tilstede i estimerne, kan vi korrigere estimerne for bias. I praksis er det dog svært at vide, hvor meget bias der er tilstede i estimerne. Det er derfor vigtigt at forstå, at relativt få års data med indkomst kan være med til at underestimere estimerne og dermed overestimere mobiliteten i populationen. Hvis de pågældende surveys indeholdte flere års data med indkomst, ville det som udgangspunkt være muligt at lave en deskriptiv undersøgelse af den intergenerationelle mobilitet uden signifikante bias.

Det samme er dog ikke nødvendigvis tilfældet med undersøgelse af de kausale sammenhænge, hvor der er flere datakrav, der skal opfyldes. Den største svaghed ved at bruge surveys til at undersøge kausale sammenhænge for den intergenerationelle mobilitet er, at der er en relativ lav fleksibilitet i forhold til en undersøgelse med registerdata. Det kræver, at der er en reform, som passer på den pågældende survey, hvilket i praksis er en sjældenhed og selv hvis der er en reform, som er kompatibel, er det ikke nødvendigvis nok til at finde de kausale sammenhæng. Der findes yderligere data-mæssige krav om, at der skal kunne opstilles kontrolgrupper og at reformen har en klar effekt. Udfordringerne ved at undersøge de kausale sammenhæng kommer også til udtryk i litteraturen, hvor forskere ofte er ”opfindsomme”, når de skal finde de kausale effekter. Det er derfor ikke nødvendigvis en fordel at have registerdata, hvis ikke man kan ”opstille” et eksperiment, som kan påvise kausaliteten. I princippet ville det ”perfekte” data være tilfældet, hvor man selv indsamler data og dermed

selv opstiller eksperimentet. I praksis er dette dog utroligt ressourcekrævende, hvilket betyder, at det som udgangspunkt ikke er en løsning på problematikken. Det må derfor være en bestræbelse for lande uden registerdata at skabe registre og dermed gøre undersøgelser af de kausale sammenhæng i den intergenerationelle mobilitet mere fleksible.

En vigtig pointe ved denne afhandlings fokus på surveydata, ligger i den ofte begrænsede adgang, som forskere på alle niveauer møder, når det kommer til såkaldt big data og herunder administrativt data. Selvom nogle forskere har adgang til de skandinaviske administrative registre, er det ikke en selvfølgelighed med registre i andre udviklede økonomier som f.eks. USA. Vi ønsker derfor at belyse den udvikling, som er sket i forskningen for intergenerationel mobilitet. Dette gøres i forhold til tilgængeligheden af data for bedre at kunne forstå den relevans, som mindre surveys kan have i dag og i lige så høj grad vigtigheden af bedre og mere centraliseret data i fremtiden.

Hamermesh (2012) har ved en gennemgang af økonomiske udgivelser fra 1963-2011 beskrevet hvordan at forskere fra midten af 1980 til i dag, er gået fra at anvende små statslige surveys til at benytte store administrative datasæt med tæt på universal dækning. Hamermesh fremhæver flere pointer ved det administrative data. Først, det administrative data indsamles og udgives næsten i real tid i mange tilfælde. Hvor det førhen tog flere år at indsamle og udgive data fra et survey, er der for offentlige statistikbanker kun et lag på måned- eller årsbasis. Den hurtige dataadgang kan eksempelvis være en fordel i evalueringen af politikker og reformer, da det tilbyder mere fleksible måder at afprøve politikker. Et andet aspekt fremhævet af Hamermesh er kvantificeringen af aktiviteter, som ikke tidligere har været muligt. Især informationer som kan hentes fra skatteregistre, velfærdsprogrammer, uddannelsesregistre, er data som giver økonomer mulighed for at undersøge områder, der ikke tidligere har været empirisk muligt, men vigtige på de økonomiske og sociologiske forskningsområder.

Strukturen i data har også enorm betydning. Administrative registre er som regel kendetegnet ved at være mere dekomponerbar end surveys. At finde måder hvorpå store ustruktureret datasæt kan organiseres bliver derfor automatisk en af hovedopgaverne når man arbejder med denne type data, men samtidig også her alle fordelene opstår. Mulighederne for at kombinere og skabe nye variable er tæt på uendelige og giver derfor en utømmelig kilde til information (Einav & Levin, 2014).

Card et al. (2010) beskriver yderligere behovet for udvidelse af dataadgange i USA ved de markant større stikprøver som administrative datasæt giver. Derudover påpeger de, at problemer med attrition bias, non-response og målefejl er langt mindre ved administrative data i forhold til surveys. USA har

hidtil været førende indenfor sociologisk og økonomisk forskning - ofte baseret på amerikanske surveys, men har i konkurrence med lande med administrative datasæt mistet denne føring. Årsagen til at USA har mistet føringen beskrives af Card et al. (2010), som påpeger problematikken for administrativt data ved de krav der stilles til fortrolighed og sikkerhed. Det er f.eks. vigtigt at følsomt og værdifuldt data ikke bliver offentligt tilgængeligt. Netop dette er i højere grad en udfordring i USA, som kendetegnes ved en mere decentraliseret stat, hvor datadelingen mellem statens organer er mere begrænset end f.eks. de skandinaviske lande. Interne regler og lovgivninger umuliggør derfor et centralt datacenter som det f.eks. er tilfældet i Danmark. Sidst nævner Card et al. (2010), at den amerikanske befolkning har en grundlæggende mistillid til ideen om centralisering, som kan give staten en art monopol over sensitivt data. Der findes derfor et opgør mellem hvad amerikanske forskere ønsker sig af dataadgange og hvad den amerikanske befolkning ønsker i forhold til centralisering af data. Det er derfor ikke muligt at merge f.eks. uddannelsesdata med indkomstdata som redskab eksempelvis til at gennemføre og opstille nye kvasi-eksperimenter og research designs (Card et al., 2010). Eftersom en lang række lande ikke indsamler og opbevarer data som det er tilfældet i de skandinaviske lande, kan surveys være den eneste kilde til brugbare data, selvom det ved f.eks. kvasi-eksperimenter har en række af begrænsninger.

En anden begrænsning ved projektet er, at vi kun undersøger det danske survey DLSY, hvilket kan gøre det svært at lave en generel konklusion. Vi formår dog til dels at imødekomme dette problem ved at sammenligne de deskriptive mål (IGE, rank-rank og transitionsmatrice) med andre undersøgelser fra de mest anvendte surveys i litteraturen. Herved får vi en mere nuanceret analyse af de generelle metodiske problematikker. Vi kender ikke til andre undersøgelser fra litteraturen, som undersøger de kausale sammenhænge med DiD i et survey og kan derfor ikke direkte sammenligne med disse. Vi kan dog stadig lave nogle generelle konklusioner ud fra vores kendskab til surveys og datakravene ved at bruge DiD. Det er f.eks. generelt problematisk at udføre en analyse med DiD, når de fleste surveys ikke indeholder flere år med indkomstdata. Vi anser ikke datamæssige mangler og udfordringer som begrænsninger i dette projekt eftersom, formålet med dette projekt er at påvise disse udfordringer. Noget tyder dog på, at det danske DLSY generelt er mere begrænset i forhold til, de surveys vi har sammenlignet med. DLSY indeholder f.eks. forskellige opgørelser for indkomst i modsætningen til andre surveys, som vi har kendskab til.

Vi foreslår, at den fremtidige forskning indenfor området fokuserer på at gøre vores resultater mere generelle ved at inddrage flere surveys. En oversigt over hvordan flere surveys ser ud forskellige

steder i verden, vil være med til at give en mere nuanceret konklusion i forhold til, hvilke problematikker der er forbundet med brug af surveys. Derudover burde fremtidig forskning også fokusere på, hvordan der kan korrigeres for de generelle bias, som ofte opstår i forbindelse med brug af surveys. Det vil f.eks. være fordelagtigt at kunne korrigere for problematikken med, at der ikke altid er flere års data for indkomst i de tilgængelige surveys. En tredje gren af fremtidig forskning kunne fokusere på forskellen mellem estimerne for intergenerationel mobilitet i lande, hvor der både er administrative data og surveys tilgængeligt.

5. Konklusion

Forskning i intergenerationel mobilitet er blevet populært inden for det sidste årti samtidigt med, at den akademiske og samfundsmæssige interesse i ulighed og heriblandt ulighed i muligheder er steget. Der findes dog en del problematikker ved at måle intergenerationel mobilitet empirisk og at måle hvad der kausalt påvirker den. Disse problematikker kan ofte tilskrives begrænsede data i mange lande. Den nyeste forskning fokuserer i højere grad på at undersøge intergenerationel mobilitet kausalt, hvilket stiller yderligere krav til det anvendte data end de traditionelle deskriptive studier. Selvom de fleste forskere i nordiske lande har adgang til administrative data, er der stadig en lang række lande, som ikke har samme data tilgængeligt. Det er derfor relevant at se på de muligheder og begrænsninger der eksisterer, når man anvender longitudinale surveys. Vi har i denne afhandling undersøgt de metodiske problematikker ved den intergenerationelle mobilitet i observationsbegrænsede longitudinale surveys. Dette gøres ud fra følgende problemformulering:

Hvilke metodiske problematikker opstår der i forbindelse med at undersøge den intergenerationelle indkomstmobilitet med data fra typiske observationsbegrænsede longitudinale surveys?

For at besvare den overordnede problemformulering har vi opstillet to underspørgsmål:

1. Hvilke metodiske problematikker opstår der i forbindelse med at estimere den intergenerationelle mobilitet deskriptivt i observationsbegrænsede surveys (IGE, rank-rank og transitionsmatricer)?
2. Hvilke metodiske problematikker opstår der i forbindelse med at finde de kausale sammenhænge mellem den intergenerationelle mobilitet og tidlig uddannelse?

Vi benytter det danske survey DLSY som case for at påvise de metodiske problematikker ved at bruge observationsbegrænsede longitudinale surveys, når man undersøger intergenerationel mobilitet. En relevans i forhold til at undersøge det danske survey DLSY er også, at der i Danmark både findes

surveydata og administrativt data, hvilket gør det muligt at sammenholde estimerer og databegrænsninger med en typisk datamæssig ideal situation.

Det første underspørgsmål besvares ved at sammenligne undersøgelser fra litteraturen som har brugt lignende data med vores undersøgelse af IGE, rank-rank korrelationer og transitionsmatricer i DLSY. Her finder vi, at det generelt er problematisk at der alene er adgang til få års indkomstdata, hvilket kan medføre en underestimering af IGE og rank-rank korrelationerne og dermed en overestimering af mobiliteten. Dette er i overensstemmelse med tidligere metodisk forskning, som har påpeget, at det er vigtigt at bruge mål for indkomsten, som er en approksimation af den permanente indkomst. Kendte problematikker i dette indebærer transitoriske shocks til indkomsten og life cycle bias. For at undgå disse problematikker anbefales det at måle et gennemsnit for indkomsten over så mange år som muligt og på et tidspunkt i individernes liv tæt på deres permanente indkomst. Vi finder yderligere, at det generelt ikke er et problem at løse problematikken ved life cycle bias, eftersom der ofte er indkomstdata for individerne, når de er mellem 30-45 år gamle. Der findes dog forskellige bud i litteraturen på, hvornår det er bedst at måle indkomsten for at undgå life cycle bias.

Generelt findes der noget usikkerhed vedrørende validiteten af svarene på spørgsmålene i surveys, som er med til at skabe en vis måleusikkerhed for målene af indkomst. Vi finder yderligere, at der kan være inkonsistens i spørgsmålene i surveys og metoderne hvorpå indkomsten er indsamlet. Dette er med til at problematisere undersøgelse og sammenligning af intergenerationel mobilitet i disse surveys.

Overordnet finder vi, at det er muligt at undersøge intergenerationel mobilitet empirisk i observationsbegrænsede longitudinale surveys, men estimererne oftest vil være præget af bias, som underestimerer målene og dermed overestimerer mobiliteten. Mazumder (2005) påviser f.eks., at tidligere IGE-estimerer fra USA er blevet underestimeret med omkring 30% og foreslår derfor at estimererne skal opskaleres med 30%.

I det andet underspørgsmål forsøger vi at påvise de metodiske problematikker ved at undersøge kausale sammenhænge ved intergenerationel mobilitet. Dette gøres ved at opstille en case med udgangspunkt i den danske reform Bistandsloven 1976, som bl.a. havde til formål at øge tidlig uddannelse, hvor vi viser problematikkerne ved at undersøge de kausale sammenhænge i et observationsbegrænset longitudinal survey med brug af den ofte anvendte Difference-in-Difference estimationsmetode. I vores gennemgang af DiD estimationsmetoden viser vi, at der er bestemte datakrav, som skal opfyldes, hvis undersøgelse med DiD skal være valide. Det er et krav, at man kan opstille en klar treatment- og

kontrolgruppe, således effekterne af den ønskede reform kan isoleres. For at gøre dette skal en række antagelser opfyldes, hvoraf den vigtigste er antagelsen om parallelle trends. Vi finder her nogle klare begrænsninger ved at anvende DLSY. Overordnet er det problematisk, at datasættet kun indeholder indkomstdata fra ét år, hvilket umuliggør at opstille parallelle trendlinjer. Dette problem vil ikke nødvendigvis eksistere i andre surveys som f.eks. PSID. Ikke desto mindre skal det anvendte data tidsmæssigt stemme med den lovændring eller anden påvirkning, man ønsker at analysere. Samtidig skal lovændringen indføres på en måde, der gør det muligt at opstille særskilte kontrol- og reformgrupper, hvilket der tilsvarende kan være begrænsede muligheder for med surveys. Som understøttende data for udfordringen i at bruge surveys til at undersøge kausale effekter på intergenerational mobilitet kan nævnes, at vi i vores litteraturgennemgang kun har fundet et enkelt studie, der anvender longitudinalt survey data til at estimere den kausale effekter af offentlige midler (Mayer & Lopoo, 2008), mens der findes adskillige eksempler, som er baserede på administrativt data, som eksempelvis Chetty & Hendren (2018), Havnes & Mogstad (2012) og Pekkarinen et al (2009).

Vores bidrag til litteraturen er at samle de forskellige problematikker vedrørende undersøgelsen af intergenerational mobilitet i observationsbegrænsede surveys. Herudover bidrager vi også til forståelsen af problemerne ved at undersøge de kausale sammenhænge, som efter vores kendskab ikke er påvist i litteraturen.

Litteraturliste

- Angrist, J.D. & Pischke, J. (2008), "Mostly Harmless Econometrics", Princeton University Press
- Baker, M. & Solon, G. (2003), "Earnings dynamics and inequality among Canadian men, 1976-1992", *Journal of Labor Economics* 21, 289–321.
- Becker, G. S. & Tomes, N. (1979), "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility", *Journal of Political Economy*, 87(6):1153–1189.
- Becker, G. S. & Tomes, N. (1986), "Human Capital and the Rise and Fall of Families", *Journal of Labor Economics*, 4(3, Part 2):S1–S39.
- Black, S. E. & Devereux, P. J. (2011), "Recent developments in intergenerational mobility", *Handbook of Labour Economics*, Chapter 16, Volume 4, Part B, Elsevier.
- Blanden, J., Goodman, A., Gregg, P. & Machin, S. (2002), "Changes in Intergenerational Mobility in Britain", CMPO Working Paper Series No. 01/43.
- Bloome, D. (2009), "Violations of SUTVA", *Social Science Statistics Blog* - Harvard University
- Bloome, D. & Western, B. (2011), "Cohort change and racial differences in educational and income mobility", *Social Forces* 90 (2): 375–95.
- Boserup, S., Kopczuk, W. & Kreiner, C. (2013), "Intergenerational Wealth Mobility: Evidence from Danish Wealth Records of Three Generations", Univ. of Copenhagen mimeo.
- Björklund, A. & Jäntti, M. (1997), "Intergenerational Income Mobility in Sweden Compared to the United States", *The American Economic Review* 87 (5): 1009–18.
- Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Jäntti, M., Eriksson, T. & Österbacka, E. (2007), "Nonlinearities in Intergenerational Earnings Mobility: Consequences for Cross Country Comparisons", *The Economic Journal*, 117:72–92.
- Bhattacharya, D. & Mazumder, B. (2011), "A nonparametric analysis of black-white differences in intergenerational income mobility in the United States", *Quantitative Economics* 2:335–79
- Card, D. & Kruger, A.B. (1993), "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania", NBER Working Paper No. 4509

- Card, D, Chetty, R, Feldstein, M & Saez, E (2010), "Expanding Access to Administrative Data for Research in the United States", SSRN Electronic Journal
- Case, A. & Paxson, C. (2006), "Children's Health and Social Mobility", *The Future of Children*. 16 (2): 151–73.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. & Yagan, D. (2011), "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR", *Quarterly Journal of Economics* 126 (4): 1593–660.
- Chetty, R., Hendren, N., Kline, P. & Saez, E. (2014), "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States", *Quarterly Journal of Economics*, 129(4):1553–1623.
- Chetty, R. & Hendren, N. (2018), "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects", *The Quarterly Journal of Economics*, Volume 133, Issue 3, August 2018, Pages 1107–1162.
- Cingano, F. (2014), "Trends in Income Inequality and its Impact on Economic Growth", OECD Social, Employment and Migration Working Papers, No. 163, OECD Publishing.
- Cunha, F., Heckman, J.J., Lochner, L. & Masterov, D.V. (2006), "Interpreting the Evidence on Life Cycle Skill Formation", *Handbook of the Economics of Education*, Volume 1, Chapter 12
- Currie, J. & Thomas, D. (2001), "Early Test Scores, School Quality and SES: Long Run Effects on Wage and Employment Outcomes", *Research in Labor Economics*, 20: 103–32.
- Dabla-Norris, E., Kochhar, K., Suphaphiphat, N., Ricka, F., & Tsounta, E. (2015), "Causes and Consequences of Income Inequality: A Global Perspective", IMF Staff Discussion Notes 15/13, International Monetary Fund.
- Einav, L & Levin, J (2014), "Economics in the age of big data", *Science* Vol 346, Issue 6210
- Elango, S., Heckman, J.J., Garcia, J.L. & Hojman, A. (2015), "Early Childhood Education", NBER Working Paper No. 21766
- EVA (2017), "Normering og gruppestørrelse og organisering", Danmarks Evalueringsinstitut
- Farbøl, R., Sørensen, E.A., Olesen, T.B. & Olesen, N.W. (2018), "Bistandsloven, 1974-1998", danmarkshistorien.dk - Aarhus Universitet

- Fertig, A. R. (2001), "Trends in Intergenerational Earnings Mobility", Working Paper No. 01-23, Center for Research on Child Wellbeing, Princeton University.
- Frisch, R. & Waugh, F. V. (1933), "Partial Time Regressions as Compared with Individual Trends", *Econometrica*, 1(4):387–401.
- FOA (2019), "Dagplejens historie", FOA - Fag og Arbejde
- Glavind, N. & Pade, S. (2018), "Daginstitutionernes hverdag 2018", Bureau 2000, FOA.
- Haider, S. & Solon, G. (2006), "Life-Cycle Variation in the Association between Current and Life-time Earnings", *The American Economic Review*, 96(4):1308–1320.
- Hamermesh, Daniel S. (2012), "Six Decades of Top Economics Publishing: Who and How?", NBER Working Paper 18635
- Havnes, T. & Mogstad, M. (2012), "Is Universal Child Care Leveling the Playing Field?", CESIFO WORKING PAPER NO. 4014
- Heckman, J.J. (2019), "The Heckman Curve", Heckman: The economics of human potential
- Heckman, J.J., Pinto, R. & Savelyev, P. (2013), "Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes", *American Economic Review* 103 (6): 2052–86.
- Holmlund, H. (2008), "Intergenerational mobility and assortative mating: effects of an educational reform", Discussion paper 0091, Centre for Economics of Education, LSE.
- Jäntti, M., Bratsberg, B., Røed, K., Raaum, O., Österbacka, E., Björklund, A., & Eriksson, T. (2006), "American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States", *IZA Discussion Paper*, (1938):1–40.
- Landersø, R. & Heckman, J. J. (2017), "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the U.S.", Study Paper no. 112, Rockwool Foundation Research Unit.
- Lechner, M. (2011), "The Estimation of Causal Effects by Difference-in-Difference Methods", *Foundations and Trends in Econometrics* Vol. 4

- Levine, D. I. & Mazumder, B. (2002), “Choosing the right parents: changes in the intergenerational transmission of inequality between 1980 and the early 1990s”, Federal Reserve Bank of Chicago Working Paper 2002-08, Chicago, IL.
- Mailman School of Public Health (2019), "Difference-in-Difference Estimation", Columbia University Mailman School of Public Health
- Mayer, S. E. & Lopoo, L. M. (2005), “Government spending and intergenerational mobility”, *Journal of Public Economics*, Volume 92, Issues 1–2, Pages 139–158.
- Mayer, S. E. & Lopoo, L. M. (2008), “Has the intergenerational transmission of economic status changed?”, *Journal of Human Resources* 40 (1): 169–85.
- Mazumder, B. (2005), “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data”, *Review of Economics and Statistics*, 87(2):235–255.
- Meghir, C. & Marten, P. (2005), “Educational reform, ability and family background”, *American Economic Review* 95, 414–424.
- Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D. G., Ramasubiah, R. G. N. & Thewissen, S. (2018), “Fair Progress? Economic Mobility across Generations around the World”, Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO.
- Nybom, M. & Stuhler, J. (2016), “Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation”, *Journal of Human Resources*, 51(1):239–268.
- O’Neill, D., Sweetman, O. & van de Gaer, D. (2007), “The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility”, *Journal of Economic Inequality* 5, 159–178.
- Pekkarinen, T., Uusitalo, R., & Kerr, S. (2009), “School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform”, *Journal of Public Economics*, 93(7-8):965–973.
- Piketty, T. (2014), “Capital In The Twenty-First Century”, The Belknap Press of Harvard University Press, Cambridge Massachusetts.

Plug, E. (2004), “Estimating the Effect of Mother's Schooling on Children's Schooling Using a Sample of Adoptees”, *American Economic Review*, 94 (1): 358-368.

Rawls, J. (1971), “A Theory of Justice”, The Belknap Press of Harvard University Press, Cambridge Massachusetts.

Reville, R. T., (1995), “Intertemporal and life cycle variation in measured intergenerational earnings mobility”, RAND working paper, Santa Monica, CA: RAND Corporation.

Sacerdote, B. (2007), “How Large are the Effects from Changes in Family Environment? A Study of Korean American Adoptees”, *The Quarterly Journal of Economics*, Volume 122, Issue 1, February 2007, Pages 119–157.

Socialministeriet (2019), "Early Childhood Education and Care, Socialministeriet

Solon, G. (2004), “A Model of Intergenerational Mobility Variation over Time and Place. Generational Income Mobility in North America and Europe”, Cambridge University Press, Cambridge, pp. 38–47.

Solon, G. (1992), “Intergenerational Income Mobility”, *American Economic Review*, (June):393–408.

Torche, F. (2015), “Analyses of Intergenerational Mobility: An Interdisciplinary Review”, *The ANNALS of the American Academy of Political and Social Science*, 657(1):37–62.

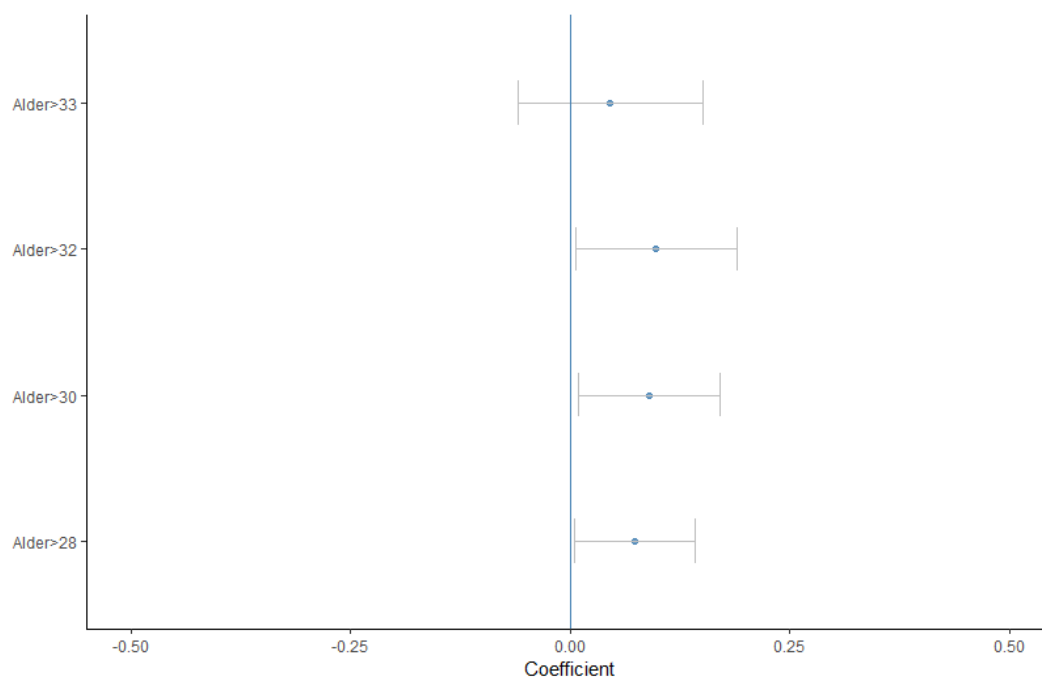
UK Data Service (2015), "Analysing change over time: repeated cross sectional and longitudinal survey data", UK Data Service

Zimmerman, D.J., (1992), “Regression toward mediocrity in economic stature”, *American Economic Review* 82, 409–429.

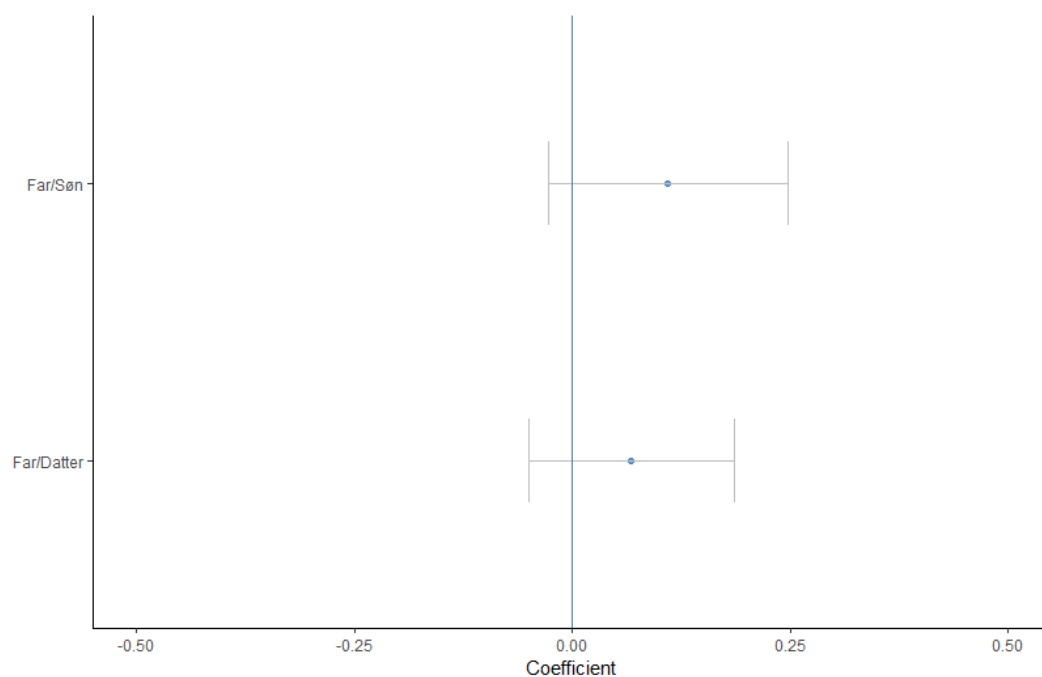
Appendiks

Appendiks A

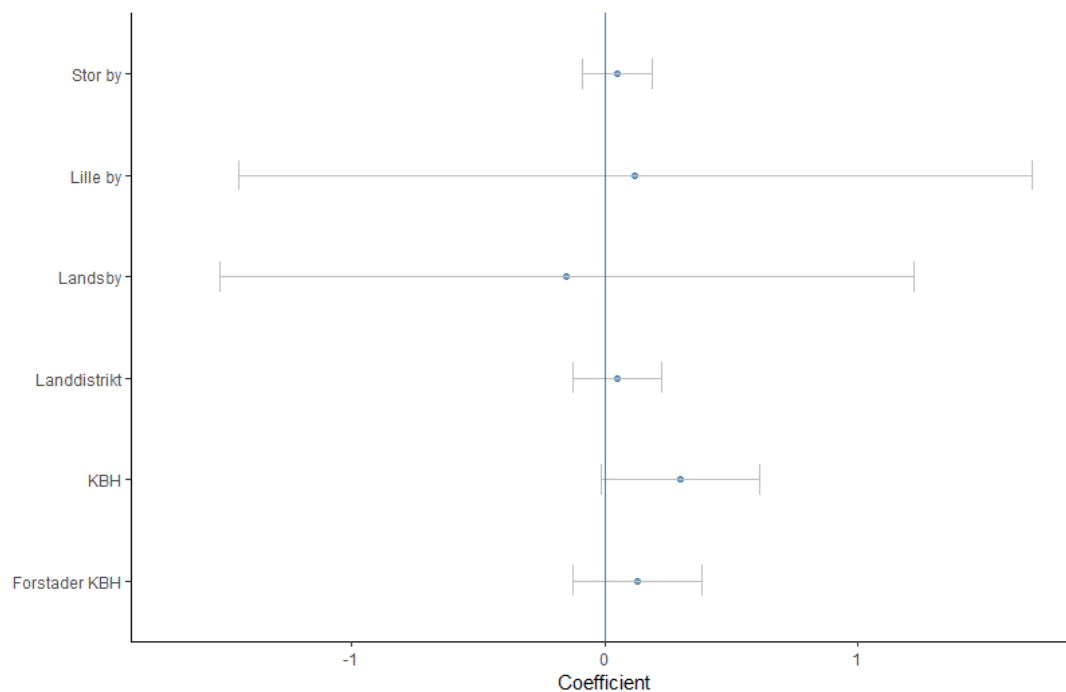
Figur A.1 – Figuren viser koefficientestimerne af IGE i forhold til standardfejlene, når der er forskellige inddelinger af alder.



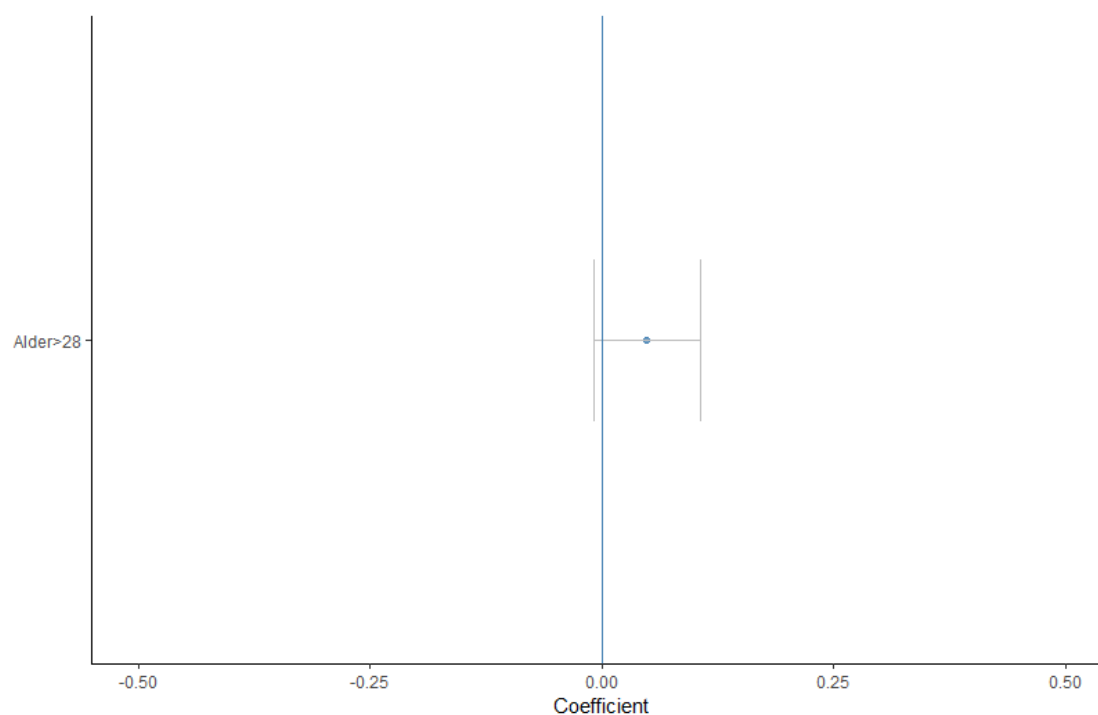
Figur A.2 – Figuren viser koefficientestimerne af IGE i forhold til standardfejlene, når der er opdelt efter far og søn og far og datter.



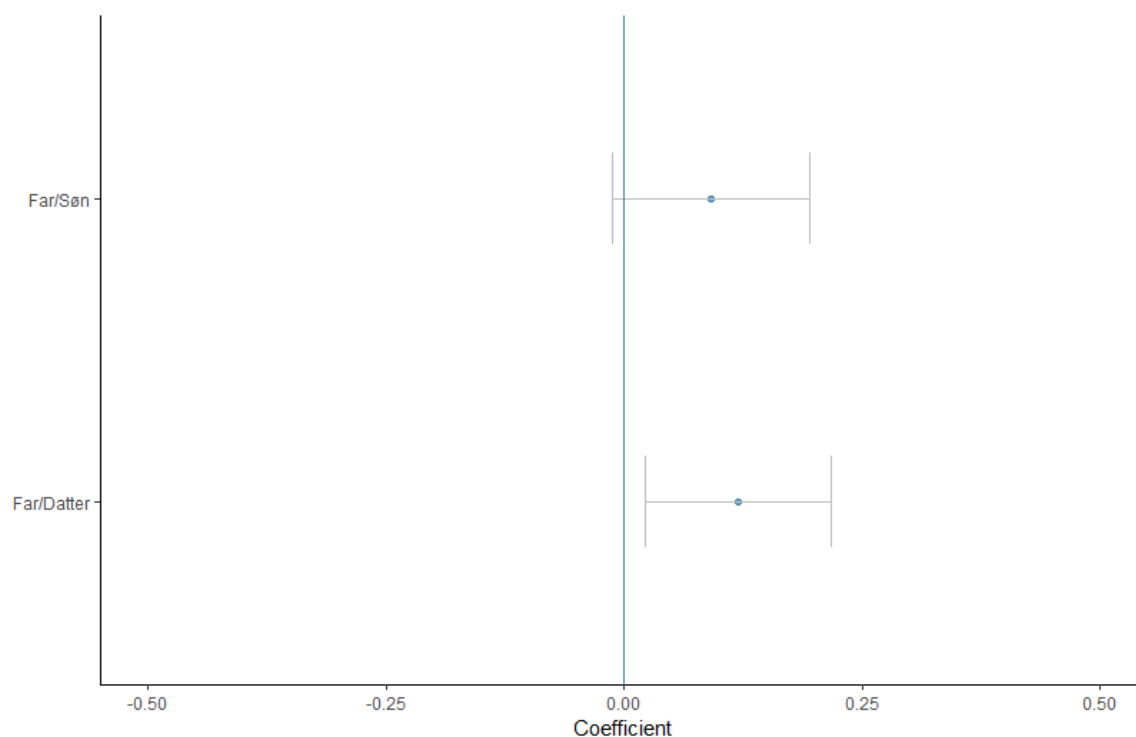
Figur A.3 - Figuren viser koefficientestimerne af IGE i forhold til standardfejlene, når der er opdelt efter geografi.



Figur A.4 – Figuren viser koefficientestimatet af rank-rank korrelationen i forhold til standardfejlen.



Figur A.5 – Figuren viser koefficientestimerne af rank-rank korrelationerne i forhold til standardfejlene, når der er opdelt efter far og søn og far og datter.



Figur A.6 - Figuren viser koefficientestimerne af rank-rank korrelationerne i forhold til standardfejlene, når der er opdelt efter geografi.

