

Baggrundsnotat22. januar 2009
Sekretariatet

Dannelse af restliv og livsforløb til belysning af flerårige fordelingsvirkninger

Overordnet set viser analyserne at:

5

- Indkomstfordelingen i de konstruerede forløb er i overensstemmelse med den faktiske indkomstfordeling.
- Indkomstmobiliteten i de konstruerede forløb ligger tæt op ad den faktiske indkomstmobilitet.
- 10 • Den isolerede ”matchfejl” er forholdsvis lille, og den beregnede Gini-koefficient for livsforløb vurderes kun at være svagt undervurderet.

I *afsnit 1* belyses baggrunden for at benytte konstruerede forløb samt metodemæssige overvejelser i forbindelse med dannelsen af disse forløb (ved statistisk match). I *afsnit 2* sammenlignes indkomstmobiliteten i de konstruerede restliv med den faktiske udvikling over en 12-årig periode. I *afsnit 3* opstilles beregningsprincipper til kvantificering af den isolerede ”matchfejl”.

15

1. Dannelse af restliv og livsforløb

20 Restlivsmodellen tager udgangspunkt i befolkningen i 2005 og konstruerer indkomster mv. frem i tiden for disse personer ved anvendelse af såkaldte statistisk match¹. Grundprincippet i dannelsen af restlivsforløbene er, at hver person i 2005 ”forlænges” ved at tage udgangspunkt i lovmodellens flerårige stikprøve i en periode på 4 år (2002-05) og anvende dette datagrundlag til at videreføre livsforløbet for den enkelte med information fra en sammenlignelig person i 2002. Alle personers skattebetaling mv. i de 4 år er beregnet efter regler for 2009, mens beløb er i 2005-niveau.

25

Videreførelsen af livsforløbet for en 40-årig i 2005 foretages fx ved at finde en 40-årig i 2002 med tilsvarende karakteristika og ”sætte den 40-årige i 2002 sammen med” den 40-årige i 2005. Derved vil der være information om den 40-årige i 2005 frem til ”2008”, hvor vedkommende vil være 43 år. Tilsvarende forlænges den 43-årige i ”2008” med en 43-årig i 2002 osv. I sammensætningen af delperioderne korrigeres for inflation og vækst. Videreførelsen af et liv ophører, når personen dør, udvandrer eller bliver 90 år.

30

¹ Det skal bemærkes, at der ved dannelse af livsforløb tages udgangspunkt i 49-51-årige personer i 2005, hvorefter livsforløbene matches opad mod 90 år og nedad mod 18 år. Livsforløbene er derfor ikke på samme måde knyttet til ”virkelige” personer, som det er tilfældet for restlivsforløbene. Livsforløbene kan til en vis grad siges at være repræsentative for de 49-51-årige med hensyn til uddannelsesniveau og enkelte andre karakteristika af mere permanent karakter.

35 For både livsforløb og restliv gælder altså, at de er konstrueret med udgangspunkt i den konjunktursituation, lovgivning, dødelighed, uddannelsesadfærd mv., som har været gældende i perioden 2002-05.

Gruppering af personer

40 For at opnå så realistiske forløb som muligt tilstræbes, at de personer, som sammenkobles, ligner hinanden mest muligt. Personerne grupperes derfor efter en række karakteristika inden matchet. Inden for hver gruppe sammenkobles personerne efter en ikke-kategoriseret oplysning – en såkaldt ”matchvariabel”. Valget af karakteristika og matchvariabel er væsentligt for de resulterende forløb og afhængig af formålet med analysen.

45 Der er i de hidtidige analyser til skattekommissionen anvendt forskellige matchgrupper afhængig af det konkrete beregningseksempel. Ved analyser af ændringer i de progressive skatter er personerne inddelt efter 6 karakteristika, mens der ved analyser af ændringer i beskatningen af henholdsvis kapitalindkomst og aktieindkomst har været anvendt 3 karakteristika, *jf. tabel 1*. Fælles for de valgte metoder er, at de alle resulterer i en opdeling på 432 grupper inden for hver alder².

50

Tabel 1. Matchkategorier ved de enkelte skatteeksperimenter						
Indkomstbeskatning			Kapitalindkomstbeskatning		Aktieindkomstbeskatning	
		Antal kat.		Antal kat.		Antal kat.
1. niveau	Køn	2	Køn	2	Køn	2
2. niveau	Uddannelse	6	Uddannelse	6	Uddannelse	6
3. niveau	Antal voksne	2	Kapitalindkomst	36	Aktieindkomst	36
4. niveau	Ejerforhold	2				
5. niveau	Antal børn	3				
6. niveau	Kapitalindk.	3				
I alt		432		432		432

Note: Initialt opdeles på alder.

55 Baggrunden for, at der opdeles mere detaljeret på kapital- og aktieindkomst (36 lige store grupper) i beregningerne af ændret kapital- og aktieindkomstbeskatning er, at det er afgørende, at de personer, der matches, har nogenlunde samme niveau for kapital- eller aktieindkomst. Opdelingen betyder fx, at en person med en negativ kapitalindkomst på 50.000 kr. vil blive matchet med en person med ca. samme negative kapitalindkomst.

Valg af matchvariabel

60 Valget af matchvariabel har betydning for hvilke sammenlignelige personer, der matches. Traditionelt har *husstandsækvivaleret disponibel indkomst* været benyttet som matchvariabel. Den disponible indkomst er beregnet som bruttoindkomsten (erhvervsindkomst, overførselsindkomst samt kapitalindkomst mv.) fratrukket personlige skatter samt fradrag for

² For personer over ca. 80 år er det ikke muligt at foretage en opdeling på uddannelse, således at der for de berørte ældre ”kun” er 72 grupper.

private pensionsindbetalinger. Den disponible indkomst er det indkomstbegreb, der danner grundlag for analyserne af indkomstforskelle, lavindkomstgruppe mv.

65 Det er således størrelsen af den disponible indkomst, der (udover matchgrupperne) danner grundlag for matchet, hvorimod sammensætningen af den disponible indkomst ikke spiller nogen rolle. Der er således risiko for, at en person, hvis indkomst består af fx lønindkomst, matches med en person, hvis indkomst består af offentlige overførsler, idet disse personer vil kunne have ens disponibel indkomst. I sammenhæng med fx en forhøjelse af
70 beskæftigelsesfradraget vil disse to personer få meget forskellige gevinster.

Brugen af ækvivaleret disponibel indkomst i matchningen kan derudover betyde, at en person med høj indkomst, som er gift med en person med lav indkomst, kan blive matchet med en person med mellemindkomst, der er gift med en anden person med mellemindkomst, idet disse to personer vil kunne have meget ens ækvivaleret disponibel indkomst. I sammenhæng med fx en nedsættelse af topskattesatsen vil de to personer få meget forskellige gevinster. Matchet på ækvivaleret indkomst kan altså risikere at mindske spredningen i gevinsten af skatteomlægningen set over restlivet.

80 Ved skatteeksperimenter i relation til beskatning af arbejdsindkomst kan det derfor være hensigtsmæssigt at anvende et andet indkomstbegreb – fx personlig indkomst eller grundlaget for arbejdsmarkedsbidraget, samt udelukkende benytte individuelle indkomstforhold – uden ækvivalering.

85 Da de skattemæssige indkomstgrundlag for fx kommuneskat, mellem- og topskat ikke er de samme, burde valget af matchvariabel i princippet sikre, at personerne, der kobles sammen, som udgangspunkt har tilnærmelsesvis samme grundlag for disse skatter. Med den anvendte metode er det imidlertid kun muligt at foretage match på en indkomstvariabel (indenfor hver matchgruppe), og der er derfor ikke umiddelbart muligt at matche på
90 flere skattegrundlag samtidigt.

2. Konstruerede restliv sammenlignet med faktisk udvikling

For at belyse forskellen mellem de faktiske (historiske) data og de matchede forløb sammenholdes den faktiske *indkomstudvikling over en 12-årig periode* med de konstruerede restlivsforløb. Konkret er der foretaget følgende analyser:

- Personer, der i udgangsåret er studerende, følges, og deres indkomstfordeling belyses 12 år senere – både med de konstruerede restlivsforløb og på baggrund af faktiske forløb (fra 1993-2005).
- 100 • Indkomstmobiliteten over 12 år belyses ved at sammenholde placeringen i indkomstfordelingen i udgangsåret med placeringen i indkomstfordelingen 12 år senere for henholdsvis konstruerede restlivsforløb og faktiske forløb.

105 Det bemærkes, at sammenligninger af den faktiske indkomstudvikling med de konstruerede forløb er påvirket af forskelle i konjunkturer mv. i 12 års perioden og i den periode,

der ligger til grund for de matchede forløb (2002-05). I afsnit 3 foretages beregninger af den isolerede ”matchfejl”.

Indkomstfordeling for studerende (1993-2005)

110 Det kan ikke forventes, at indkomstfordelingen er præcis den samme i de to opgørelsesmetoder. Det skyldes blandt andet, at restlivene er konstrueret på baggrund af forløbsdata for perioden 2002-05, mens de faktiske forløb er baseret på 1993-2005, hvorved forskelle i konjunkturer, studieadfærd mv. i de to perioder vil påvirke resultaterne.

115 Der er forholdsvis stor overensstemmelse i studerendes indkomstfordeling efter 12 år, når restlivsforløbene sammenlignes med den faktiske udvikling fra 1993 til 2005, *jf. tabel 2*. Af dem, som er studerende i udgangsåret (1993), er 7,7 pct. placeret i 1. indkomstdecil 12 år senere. For restlivsmodellen er andelen af studerende placeret i 1. indkomstdecil 12 år senere 8,8 pct. – dvs. 1,1 pct.-enheder højere. Generelt er en lidt større andel af de studerende placeret i de lavere indkomstdeciler ved anvendelse af restlivsmodellen, og en lidt mindre andel placeret i de højeste indkomstgrupper end ved beregninger baseret på de historiske registeroplysninger. Samlet set er der imidlertid et forholdsvis pænt sammenfald i de to fordelinger. Den højere andel studerende i de lavere indkomstdeciler kan skyldes, at restlivsmodellen ikke kontrollerer for studietiden. Variationen i studietiden er overvurderet i de konstruerede restliv, og derfor er der ”for mange”, der fortsat er studerende 12 år efter. Det trækker isoleret set i retning af, at ”for mange” er i 1. indkomstdecil i slutåret.

Tabel 2. Fordeling af studerende efter 12 år

	Indkomstdecil									
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Register (1993-2005)	7,7	5,3	7,9	10,6	12,3	12,9	12,4	11,6	10,5	8,8
Restliv (2005-2017)	8,8	6,0	9,2	11,7	12,8	12,8	11,8	10,4	9,1	7,2
Forskel	1,1	0,7	1,3	1,1	0,6	-0,1	-0,6	-1,1	-1,3	-1,6

Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen i både startåret og i slutåret. Studerende i startåret er udvalgt. Kun personer over 17 år i udgangsåret og under 91 år i slutåret er medtaget. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.
Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

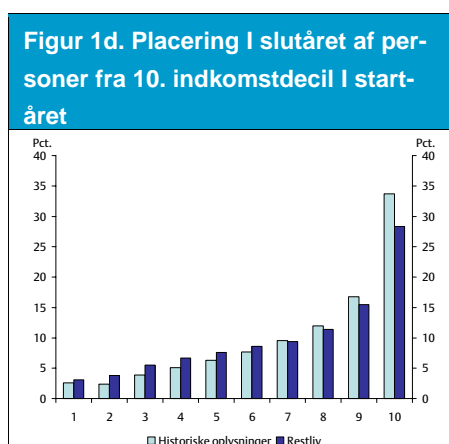
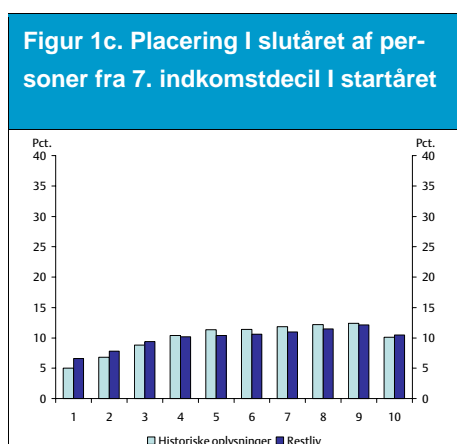
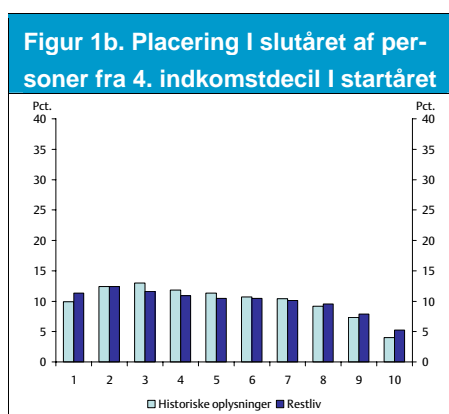
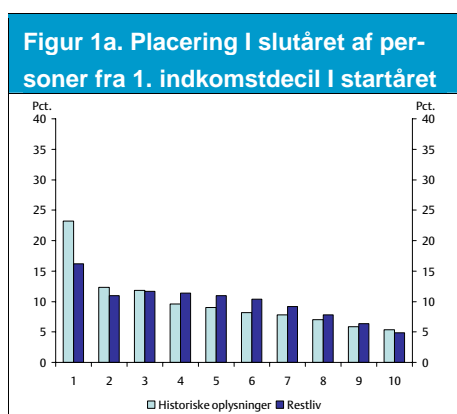
Indkomstmobilitet (1993-2005)

130 Indkomstmobiliteten i de konstruerede restliv er nedenfor sammenlignet med den faktiske indkomstmobilitet efter samme principper som i opgørelsen ovenfor. Også her er der forholdsvis stor overensstemmelse mellem restlivsforløbene og den faktiske udvikling, *jf. figur 1a-1d*.

135 Af personer, der i startåret var placeret i 1. indkomstdecil, er godt 23 pct. også placeret i 1. indkomstdecil 12 år senere (baseret på perioden 1993-2005), *jf. figur 1a*. Opgjort på bag-

grund af restlivsmodellen er denne andel godt 16 pct. Der er således en forskel mellem de historiske registeroplysninger og restlivsmodellen på omtrent 7 pct.-enheder.

- 140 For personer, der i udgangsåret er placeret i 10. indkomstdecil, gør tilsvarende sig gældende. Fra de historiske registeroplysninger er omtrent en tredjedel af gruppen placeret i 10. indkomstdecil, mens andelen ved anvendelse af restlivsmodellen er godt 28 pct., *jf. figur 1d.*
- 145 For personer, der i startåret var placeret i midten af indkomstfordelingen, er der tæt ved fuldstændig overensstemmelse mellem indkomstmobiliteten i restlivsmodellen og indkomstmobiliteten i de historiske registeroplysninger, *jf. figur 1b og 1c.*



- 150 Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen i både startåret og i slutåret. Kun personer over 17 år i udgangsåret og under 91 i slutåret er medtaget. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau. Kilde: Lovmodelberegninger på basis af stikprover på 33,3 pct. af befolkningen.

- 155 Modellen til konstruktion af restlivene har en tendens til at trække folk mod midten af ”indkomstfordelingen”. Det skyldes den tekniske måde, restlivene er dannet på med statistiske match. Ved hvert match er der en sandsynlighed for, at personen videreføres med en person med andre bagvedliggende karakteristika, selvom der tages højde for en række baggrundsfaktorer i matchet, herunder indkomst, uddannelse, køn, antal børn, ejer/lejer mv. Den ”sammenblanding”, der således sker ved hvert match bidrager dermed til at re-
- 160

ducere sandsynligheden for, at en person permanent/længerevarende er placeret i de højeste eller laveste indkomst-grupper, *jf. bilag 1*.

3. Kvantificering af ”matchfejl” ved dannelse af restliv/livsforløb

165 Sammenligningen ovenfor mellem de konstruerede forløb og den historiske udvikling er som nævnt påvirket af konjunkturer, uddannelsesadfærd mv., og forskellene kan således ikke udelukkende tilskrives ”matchfejlen”. For at kunne belyse den isolerede virkning af ”matchfejlen” er der foretaget analyser, hvor sammenhængende oplysninger for personer i perioden 2002-05 sammenholdes med forløb konstrueret ved hjælp af statistisk match på baggrund af oplysninger fra samme periode. I denne sammenhæng er foretaget følgende analyser:

- Indkomstmobiliteten i den betragtede periode belyses ved at sammenligne placeringen i indkomstfordelingen i 2002 med placeringen i indkomstfordelingen i 2005 for henholdsvis konstruerede forløb i en 4-årig periode og faktisk forløb.
- Forskelle i fordelingen af gennemsnitlige indkomster for 2002-2005 mellem de konstruerede forløb og de faktiske forløb belyses.
- Der laves en følsomhedsanalyse med hensyn til grupperingen af personer, der indgår i den statistiske match.
- 180 • De statistiske egenskaber for de konstruerede forløb sammenlignes med de statistiske egenskaber ved de faktiske forløb.

For at teste den *isolerede effekt* af at introducere ét match sammenholdes den faktiske indkomstudviklingen for personer fra 2002-05 med indkomstudviklingen for konstruerede indkomstforløb fra 2002-05 baseret på et enkelt match i 2003 – dvs. oplysninger fra 2002-03 for en person forlænges (matches) med oplysninger fra en ”lignende” person fra 2004-05. De konstruerede 4-årige forløb er således dannet ved at sammenkoble to sammenlignelige personer i midteråret 2003.

190 For at undgå, at danne de originale 4-årige indkomstforløb ved matchet, deles populationen op i to lige store dele. Den ene halvdel benyttes i perioden 2002-03, og den anden halvdel benyttes i perioden 2004-05. Der matches på oplysninger fra 2003, og oplysningerne fra første sæt personer (2002-03) bibeholdes for året.

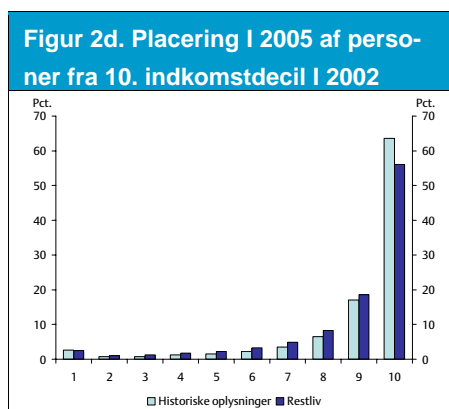
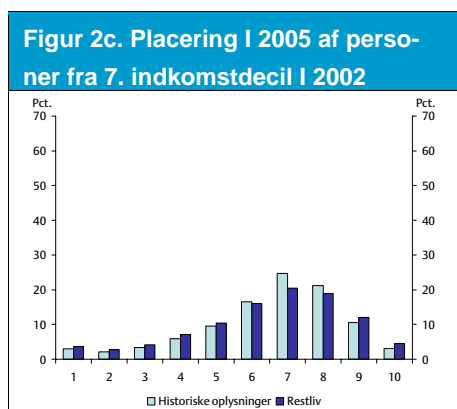
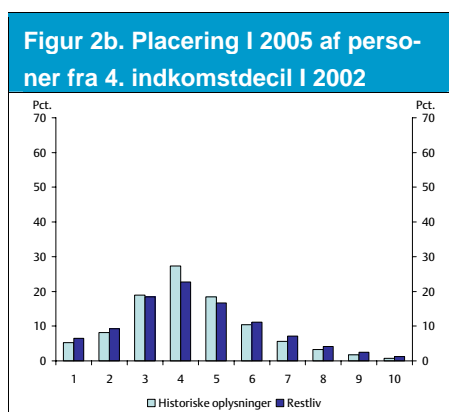
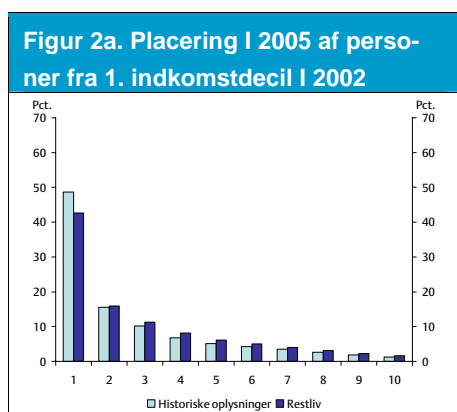
195 Ved sammenligning med den faktiske 4-årige indkomstudvikling tages der udgangspunkt i præcis samme ”halvdel”, som benyttes i de matchede forløb fra 2002-03. Det er således kun indkomstoplysningerne for 2004 og 2005, der er forskellige i de to ”sæt” af 4-årige indkomstforløb.

200 For at underbygge resultaterne af analysen af matchfejls betydning for indkomstfordelingen dannes tilsvarende forløb for perioden 2002-05 baseret på 3 match, dvs. match i 2002, 2003 og i 2004.

Indkomstmobilitet

205 Ved introduktion af én match påvirkes indkomstmobiliteten gennem perioden ikke væsentligt. Der er dog en tendens til, at færre personer i de matchede forløb bibeholder deres decilplacering i indkomstfordelingen, end det er tilfældet i de faktiske forløb, *jf. figur 2a-2d.*

210 Dette resultat er udtryk for, at der ved hvert match ”tilføjes” for meget indkomstmobilitet, da der ved hvert match sker en sammenblanding af personer med forskellige karakteristika – selvom det så vidt muligt sikres, at det er sammenlignelige personer, der matches.



215 Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen og panelet i både startåret og i slutåret. Kun personer, der findes i 2002 i begge populationer indgår i figuren. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.

Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

220 Sammenholdt med den faktiske indkomstmobilitet har personer i 1. og 2. indkomstdecil i de matchede forløb en lidt for stor sandsynlighed for at rykke op i indkomstfordelingen over de 4 år. Tilsvarende har personer i 9. og 10. indkomstdecil i de matchede forløb en lidt for stor tendens til at rykke ned i indkomstfordelingen hen over perioden, *jf. tabel B2.1 i bilag 2.*

225 Overvurderingen af indkomstmobiliteten indebærer, at spredningen i den gennemsnitlige indkomst over en længere periode vil blive undervurderet i de konstruerede forløb. Det skyldes blandt andet, at sandsynligheden for at befinde sig permanent/længerevarende i høj- eller lavindkomstgrupperne undervurderes. Der er med andre ord større grad af

230 ”mean-reversion” end i de faktiske forløb. Målt på fx Gini-koefficienten er det imidlertid vurderingen, at denne effekt har forholdsvis begrænset betydning, *jf. nedenfor*.

Gini-koefficient

235 Ved at tage udgangspunkt i den faktiske indkomstudvikling fra 2002-05 dannes gennemsnitlige årlige indkomster. Foretages en tilsvarende beregning på baggrund af de dannede 4-årige forløb med ét match, reduceres indkomstforskellene – beregnet ved Gini-koefficienten – med 0,221 pct.-enheder, *jf. tabel 3*.

Tabel 3. Afvigelse i Gini-koefficient på gennemsnitlige indkomster for perioden 2002-05

Afvigelse fra panel	1 match	3 match
Pct.-enheder	-0,221	-0,317
Pct.	-1,057	-1,519

Anm.: Gini-koefficienterne er dannet på baggrund af personer, som befinder sig i registret/forløbet i hele den 4-årige periode. Kun personer, der findes i 2002 i både panel og forløb indgår i tabellen. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.
Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

240 Den isolerede effekt af et enkelt match på 0,221 pct.-enheder svarer til en reduktion i Gini-koefficienten på 1,057 pct. Matches i stedet 3 gange inden for perioden reduceres Gini-koefficienten med 1,519 pct.

Sammensættes et livsforløb fra alderen 18 år til 90 år af sammenhængende oplysninger fra enkeltpersoner dækkende 3 år (4-periode match), kræves minimum 24 match.

245 Matchfejlen er som udgangspunkt et udtryk for, at personer med forskellige evner/kompetencer sammenkobles. En del af de personer, der har en positiv matchfejl ved første match, vil have en negativ matchfejl ved match nr. 2 (eller senere). Når der foretages flere match (som ved dannelsen af restliv og livsforløb), vil der således med andre ord optræde modsatrettede matchfejl, som (delvis) neutraliserer hinanden. Ved opregning af
250 matchfejlen for konstruktionen af hele livsforløb forekommer det derfor mest retvisende at betragte matchfejlen som udtryk for, at der er ”fejl” i matchet af den underliggende evnefordeling.

255 Antages at ”parvise” matchfejl kan udtrykkes som en normalfordeling ($e^{evne} = N(0, \sigma^2)$), kan den samlede fejl over livsforløb udtrykkes som summen af 24 identiske og uafhængige normalfordelte afvigelser – såfremt matchfejlene er (statistisk) uafhængige – dvs. $e_{24}^{evne} = N(0, 24 \cdot \sigma^2)$.

260 Idet standardafvigelsen på et enkelt match er normeret til én, medfører 24 match en relativ forøgelse af standardafvigelsen på 4,9 ($= \sqrt{24}$).³

³ Standardafvigelsen beregnes som kvadratroden af antallet af match.

Modellen med 3 match inden for perioden understøtter ovenstående antagelse. Ifølge antagelsen indebærer 3 match, at ”matchfejlen” øges med en faktor på 1,7 ($=\sqrt{3}$). I modellen med 3 match observeres en matchfejl på 1,519 pct., *jf. tabel 5*. Matchfejlen fra modellen med ét match på 1,057 pct. er dermed relativt set forøget med en faktor 1,437.

Hvis overstående forståelsesramme lægges til grund, kan det forventes, at afvigelsen på Gini-koefficienten ved 24 match vil være 4,9 gange afvigelsen ved ét match, dvs. 5,178 pct. En ”sand” Gini-koefficient på 14 vil dermed i de matchede livsforløb blive beregnet til 13,28.

I et eksperiment, der øger den ”sande” Gini-koefficient for livsforløbene med 0,5 pct.-enheder, vil den beregnede ændring fra de dannede livsforløb være på ca. 0,474 pct.-enheder.

275 *Følsomhedsanalyse af grupperingen af personer*

Når der dannes (standard-) livs-/restlivsforløb, opdeles personerne i hver aldersklasse i 432 grupper baseret på seks karakteristika: Køn(2), uddannelse(6), antal voksne(2), ejer/lejer(2), antal børn(3) og kapitalindkomst(3).

280 Antallet af grupper samt det konkrete valg af karakteristika, der indgår i grupperingen, påvirker som udgangspunkt størrelsen af ”matchfejlen” belyst ovenfor. Med henblik på at belyse ”matchfejls” følsomhed overfor grupperingen dannes de matchede forløb for perioden 2002-05 7 gange med forskellige detaljeringsgrader for grupperingen. Først matches direkte på alder for herefter successivt at tilføje en opdeling på yderligere et karakteristika. For hvert af disse konstruerede forløb beregnes afvigelsen på Gini-koefficienten for perioden, *jf. tabel 4*.

Tabel 4. Afvigelse i 4-årig Gini-koefficient (succesiv gruppering af personer)

Grupperingskategori	Afvigelse, pct.-enheder
Alder	-0,135
+Køn	-0,132
+Uddannelse	-0,120
+Voksne	-0,129
+Ejer/lejer	-0,136
+Børn	-0,154
+Kapitalindkomst	-0,221

Anm.: Gini-koefficienterne er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen og panelet i hele den 4-årige periode. Kun personer, der findes i 2002 i begge populationer indgår i tabellen. Personerne er matchet på ækvivaleret disponibel indkomst
 Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

290 De matchede forløb giver ved opdeling på 2 karakteristika – *køn og uddannelse* – indenfor hver aldersklasse den 4-årig indkomstfordeling, der ligger tættest op ad indkomstfordelingen i de faktiske indkomstforløb (panelet). Med disse 2 grupperingsvariable er Gini-koefficienten på de matchede forløb 0,120 pct.-enheder lavere end i de faktiske indkomst-

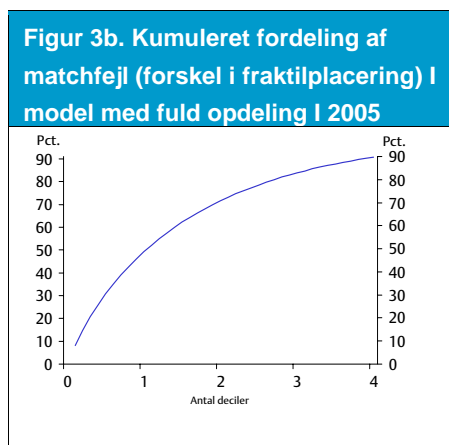
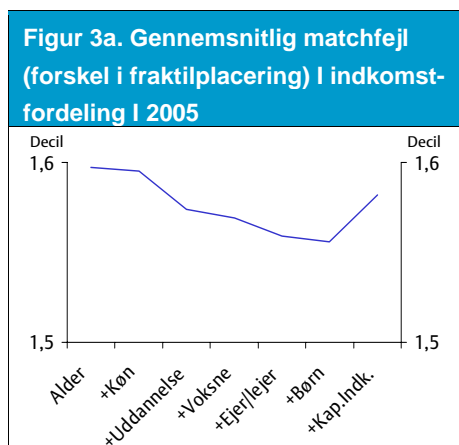
forløb. Generelt er afvigelserne fra den faktiske Gini-koefficient små med værdier omkring -0,120 til -0,150 pct.-enheder. Medtages kapitalindkomst i grupperingen, øges denne forskel til 0,221 pct.-enheder.

Variationen i de disponible indkomster afhænger af *målbare karakteristika* som fx uddannelse, familieforhold mm. samt af *ikke-målbare karakteristika* som fx intelligens, drive mm. Det antages, at jo flere karakteristika personerne inddeles efter, desto mere af variansen fra målbare karakteristika opfanges ved matchet. Givet denne opdeling antages variationen i indkomsterne indenfor gruppen at være udtryk for variationen fra ikke-målbare karakteristika.

Ved successiv opdeling af personerne vil den samlede ”matchfejl” til et vist punkt blive reduceret, men vil herefter blive forøget ved inddragelse af yderligere grupperingsvariable. Det skyldes, at den yderligere forklaring af indkomstvariationen bliver forholdsvis mindre end den samtidige forringelse af matchet på indkomstvariablen indenfor grupperne pga. færre personer i de enkelte matchgrupper.

Et andet mål for de matchede forløbs følsomhed overfor valget af gruppering er, hvor langt personerne i de matchede forløb er fra deres virkelige placering i indkomstfordelingen i 2005, jf. figur 3b. For at udtrykke denne afstand i ét tal for hele den betragtede population beregnes den gennemsnitlige numeriske forskel i fraktilplaceringen mellem den matchede indkomst og den faktiske indkomst i slutåret (2005) – dvs.

$$\frac{1}{N} \sum_{i=1}^N |F_i^{Panel} - F_i^{Match}|, \text{ hvor } F_i = \frac{n_i}{N}, \text{ jf. figur 3a.}$$



Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen og panelet i både startåret og i slutåret. Kun personer, der findes i 2002 i begge populationer indgår i tabellen. Personerne er matchet på ækvivaleret disponibel indkomst. Figuren er begrænset til at vise afstanden for knap 91 pct. af de matchede forløb.

Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

Figur 3a viser, at hvis det ønskes at minimere de matchede forløbs afstand til den virkelige placering i indkomstfordelingen i 2005, bør der opdeles på fem karakteristika – køn, uddannelse, antal voksne, ejer/lejer og antal børn – indenfor hver aldersklasse.

Fordelingen af de individuelle afstande for modellen med opdeling på alle 6 karakteristika er vist i *figur 3b*. Knap 50 pct. af de matchede forløb ender under ét decil forskudt i indkomstfordelingen i 2005.

330 Det bemærkes, at der ovenfor ikke er foretaget en systematisk analyse med henblik på at udpege de grupperingsvariable, der bidrager mest til at reducere ”matchfejlen” – enten målt ved Gini-koefficienten eller ved den gennemsnitlige fejlplacering i indkomstfordelingen.

Statistiske egenskaber

335 De gennemsnitlige disponible indkomster fordelt på deciler i 2005 for de matchede forløb ligger tæt op af de gennemsnitlige disponible indkomster i 2005 i de faktiske forløb, *jf. tabel 5*.

340 Der er dog en tendens til, at de gennemsnitlige disponible indkomster i 2005 i de matchede forløb samler sig mere omkring det samlede gennemsnit, end tilfældet er for panelet. Målt i forhold til de faktiske gennemsnitsindkomster i 2005 er gennemsnittet i de matchede forløb lidt højere for de personer, der i udgangsåret er placeret i 1.-7. indkomstdecil, mens gennemsnittet er lidt lavere for personer placeret i 8.-10. indkomstdecil i 2002. Det skyldes som tidligere nævnt, at der i de matchede forløb er en større grad af ”mean-reversion” end i de faktiske indkomstforløb.

345

Tabel 5. Disponibel indkomst i 2005, gennemsnit og spredning fordelt på deciler

Indkomstdecil 2002	Gennemsnit			Standardafvigelse		
	Panel	Forløb	Forskel	Panel	Forløb	Forskel
	Kr.	Pct.		Kr.	Pct.	
1. indkomstdecil	117.999	124.309	5,35	123.069	75.256	-38,85
2. indkomstdecil	125.446	128.982	2,82	37.166	47.887	28,85
3. indkomstdecil	140.045	142.661	1,87	40.086	48.034	19,83
4. indkomstdecil	154.410	156.796	1,55	39.142	50.693	29,51
5. indkomstdecil	168.741	170.760	1,20	42.035	64.811	54,19
6. indkomstdecil	182.932	183.568	0,35	52.836	78.107	47,83
7. indkomstdecil	197.737	197.845	0,05	51.732	83.345	61,11
8. indkomstdecil	215.794	214.883	-0,42	87.390	72.150	-17,44
9. indkomstdecil	241.146	238.666	-1,03	79.921	92.980	16,34
10. indkomstdecil	326.474	310.830	-4,79	271.028	236.300	-12,81
I alt	187.072	186.930	-0,08	82.441	84.956	3,05

Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen og panelet i både startåret og i slutåret. Kun personer, der findes i 2002 i begge populationer indgår i tabellen. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.
Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

Fordelingen af disponible indkomster i 2005 på beskæftigelsesstatus i 2002 viser i store træk samme tendens, *jf. tabel 6*. Gennemsnittet er lidt lavere for grupper, der i udgangssi-

350 tuationen har en høj indkomst, som fx ”fuldt beskæftigede”, mens gennemsnittet generelt er højere for grupper med lavindkomst i udgangssituationen, som fx ”fuldt ledige”.

Table 6. Disponibel indkomst i 2005, gennemsnit og spredning fordelt på beskæftigelsesstatus

Beskæftigelsesstatus 2002	Gennemsnit			Standardafvigelse		
	Panel	Forløb	Forskel	Panel	Forløb	Forskel
	Kr.	Kr.	Pct.	Kr.	Kr.	Pct.
Fuldt beskæftigede	218.831	216.676	-0,98	110.276	114.678	3,99
Selvstændige	199.019	198.169	-0,43	225.704	165.666	-26,60
Deltidsbeskæftigelse	190.051	188.992	-0,56	144.843	102.243	-29,41
Delårsbeskæftigelse	165.966	169.757	2,28	77.833	70.114	-9,92
Fuldt ledige m.fl.	131.587	139.744	6,20	72.123	69.721	-3,33
Uddannelsessøgende	148.737	147.885	-0,57	77.951	77.338	-0,79
Førtidspensionister	158.650	162.040	2,14	62.821	89.021	41,71
Efterlønsmodtagere	167.215	169.556	1,40	128.888	78.297	-39,25
Andre 18-66-årige	171.752	172.263	0,30	204.461	183.265	-10,37
Over 66-årige	153.287	152.875	-0,27	121.424	111.394	-8,26
I alt	187.072	186.930	-0,08	82.441	84.956	3,05

Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen og panelet i både startåret og i slutåret. Kun personer, der findes i 2002 i begge populationer indgår i tabellen. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.
Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

355

360

365

Bilag 1

Tabel B1.1. Decilfordeling i 2005 for alle I 1993 (historiske registeroplysninger)											
Decilplacering 2005											
Decilplacering	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	Alle
1993											
1. indkomstdecil	23,2	12,3	11,8	9,6	9,0	8,2	7,8	7,0	5,9	5,4	100
2. indkomstdecil	24,0	20,1	13,1	9,9	8,2	7,0	6,2	5,0	3,7	2,6	100
3. indkomstdecil	14,7	18,0	13,7	11,6	10,1	9,2	8,1	6,8	4,9	2,8	100
4. indkomstdecil	9,9	12,4	13,0	11,8	11,3	10,7	10,4	9,2	7,3	4,0	100
5. indkomstdecil	7,4	10,0	11,4	11,8	11,5	11,4	11,0	10,9	9,3	5,2	100
6. indkomstdecil	5,9	8,0	9,9	11,4	11,5	11,7	11,6	11,5	11,1	7,6	100
7. indkomstdecil	5,0	6,8	8,8	10,4	11,3	11,4	11,8	12,2	12,4	10,1	100
8. indkomstdecil	4,1	5,7	8,0	9,9	11,0	11,5	11,8	12,3	13,3	12,4	100
9. indkomstdecil	3,3	4,3	6,5	8,5	9,8	11,1	11,9	13,3	15,3	16,1	100
10. indkomstdecil	2,6	2,4	3,9	5,1	6,3	7,7	9,5	12,0	16,8	33,7	100
Alle	100	100	100	100	100	100	100	100	100	100	100

Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen i både startåret og i slutåret. Studerende i startåret er udvalgt. Kun personer over 17 år i udgangsåret og under 91 år i slutåret er medtaget. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.

Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

Tabel B1.2. Decilfordeling i 2005 og 2017 (restlivsmodellen)											
Decilplacering 2017											
Decilplacering	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	Alle
2005											
1. indkomstdecil	16,2	11,0	11,7	11,4	11,0	10,4	9,2	7,8	6,4	4,9	100
2. indkomstdecil	20,7	17,4	12,0	10,0	9,0	8,3	7,6	6,5	5,0	3,5	100
3. indkomstdecil	15,3	15,9	12,2	10,6	10,0	9,2	8,8	7,7	6,2	4,1	100
4. indkomstdecil	11,3	12,4	11,6	10,9	10,5	10,5	10,1	9,5	7,9	5,2	100
5. indkomstdecil	9,0	10,1	10,6	10,7	10,6	10,7	11,0	10,8	9,8	6,6	100
6. indkomstdecil	7,5	8,8	10,0	10,4	10,5	10,8	11,1	11,3	10,9	8,6	100
7. indkomstdecil	6,6	7,8	9,4	10,2	10,4	10,6	11,0	11,5	12,1	10,5	100
8. indkomstdecil	5,6	6,9	9,0	9,8	10,4	10,5	11,1	11,5	12,4	12,6	100
9. indkomstdecil	4,6	5,9	7,9	9,2	10,0	10,3	10,7	11,8	13,6	15,9	100
10. indkomstdecil	3,1	3,8	5,5	6,7	7,6	8,6	9,4	11,4	15,5	28,3	100
Alle	100	100	100	100	100	100	100	100	100	100	100

Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen i både startåret og i slutåret. Studerende i startåret er udvalgt. Kun personer over 17 år i udgangsåret og under 91 år i slutåret er medtaget. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.

Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

Bilag 2

Tabel B2.1. Ændring i decilplaceringer – faktisk udvikling, 2005										
Decilplacering 2005										
Decilplacering 2002	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. indkomstdecil	48,7	15,5	10,1	6,8	5,2	4,3	3,5	2,6	1,9	1,3
2. indkomstdecil	19,6	42,8	16,5	8,1	4,7	3,2	2,2	1,5	1,0	0,5
3. indkomstdecil	8,4	20,6	32,7	16,9	9,0	5,3	3,3	2,1	1,2	0,6
4. indkomstdecil	5,3	8,1	19,0	27,4	18,4	10,4	5,6	3,2	1,8	0,8
5. indkomstdecil	3,9	4,4	8,6	18,2	24,8	19,2	11,1	5,7	2,9	1,1
6. indkomstdecil	3,1	3,0	5,1	9,5	17,5	23,8	20,3	11,2	4,9	1,7
7. indkomstdecil	3,0	2,1	3,4	5,9	9,5	16,6	24,7	21,2	10,6	3,1
8. indkomstdecil	2,8	1,6	2,3	3,7	5,9	9,6	16,9	28,0	22,3	6,7
9. indkomstdecil	2,7	1,2	1,5	2,3	3,5	5,3	8,9	17,9	36,3	20,4
10. indkomstdecil	2,6	0,8	0,8	1,2	1,5	2,3	3,5	6,5	17,1	63,6

Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen i både startåret og i slutåret. Studerende i startåret er udvalgt. Kun personer over 17 år i udgangsåret og under 91 år i slutåret er medtaget. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.

Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.

Tabel B2.2. Ændring i decilplaceringer – matchede forløb, 2005										
Decilplacering 2005										
Decilplacering 2002	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. indkomstdecil	42,6	15,9	11,3	8,1	6,1	5,0	4,0	3,1	2,3	1,6
2. indkomstdecil	20,0	38,3	16,1	8,8	5,8	4,0	2,9	2,0	1,4	0,7
3. indkomstdecil	9,7	20,5	28,2	15,7	9,9	6,3	4,2	2,8	1,8	0,9
4. indkomstdecil	6,5	9,3	18,4	22,7	16,7	11,2	7,1	4,2	2,5	1,2
5. indkomstdecil	4,9	5,3	9,6	17,4	20,5	17,5	11,9	7,1	4,0	1,8
6. indkomstdecil	3,9	3,6	6,2	10,5	16,6	20,0	18,0	12,2	6,5	2,5
7. indkomstdecil	3,7	2,7	4,2	7,1	10,4	16,1	20,4	19,0	12,0	4,5
8. indkomstdecil	3,2	1,9	2,9	4,7	7,3	10,2	16,5	23,4	20,8	9,1
9. indkomstdecil	2,9	1,5	2,0	3,2	4,2	6,5	10,2	17,9	30,2	21,4
10. indkomstdecil	2,5	1,0	1,3	1,7	2,3	3,2	4,9	8,3	18,6	56,1

Anm.: Decilinddelingen er dannet på baggrund af personer, som befinder sig i registret/restlivs-modellen i både startåret og i slutåret. Studerende i startåret er udvalgt. Kun personer over 17 år i udgangsåret og under 91 år i slutåret er medtaget. Personerne er matchet på ækvivaleret disponibel indkomst efter opdeling på køn, uddannelse, antal voksne, ejer/lejer, antal børn og kapitalindkomstniveau.

Kilde: Lovmodelberegninger på basis af stikprøver på 33,3 pct. af befolkningen.