

Revision af EU-SILC

Til brugere af SILC

Dette papir beskriver revisionen af SILC. Den gennemføres i sommeren 2015 på SILC 2014 data. Når det er muligt, vil revisionen blive gennemført for tidligere år for at sikre en konsistent dataserie også tilbage i tid.

Formål med revisionen For at sikre at SILC stikprøven matcher befolkningen bedst muligt, tildeles alle undersøgelsens deltagere vægte. Disse vægte sikrer, at stikprøvens deltagere afspejler befolkningen, når det handler om alder og indkomstforhold. Det primære formål med revisionen er at begrænse variationen i disse vægte. Stor variation i vægtene giver anledning til stor usikkerhed på indikatorer, når kun en mindre del af stikprøven betragtes fx specifikke aldersgrupper.

SILC bruges i Danmark primært til internationale sammenligninger af indkomstforhold. Ved rene nationale betragtninger anbefales brug af nationale registre i stedet. Med det som udgangspunkt er det et selvstændigt mål, at forbedre sammenligneligheden på SILC i forhold til de øvrige EU-lande.

Endelig ønskes det at implementere elementer fra den revision af den nationale indkomststatistik og nationalregnskabet, som gennemførtes i foregående år.

1. Revisionens hovedelementer

1.1 Variation i vægtene

Vægt introduktion SILC er baseret på en stikprøve. En husstand bliver bedt om at deltage i fire på hinanden følgende år. I Danmark deltager hvert år omkring 5.500 ud af omkring 13.000 udtrukne husstande i undersøgelsen. I Danmark er der ca. 2,6 mio. husstande. Hver deltagende husstand i undersøgelsen repræsenterer altså omkring 500 danske husstande. Blandt andet fordi bortfaldet i undersøgelsen er skævt, får nogle husstande en lavere vægt og andre end højere vægt i kalibreringen.

Nye kvalitetskrav Danmark har meget stor variation på de vægte, som sikrer at stikprøven matcher den fulde befolkning. Eurostat indfører fra 2018 nye kvalitetskrav til SILC bl.a. med henblik på at introducere en regionalisering af SILC. Her er det primære fokus i Danmark at begrænse den teoretiske stikprøveusikkerhed.

På de væsentligste indkomstindikatorer og demografien sikrer den nuværende danske kalibrering, at SILC nøjagtigt rammer niveauerne, som kan beregnes med det fulde indkomstregister og befolkningsregistret, når der måles på landsplan. Her er altså ingen usikkerhed.

Stor variation på vægte gør resultaterne mere usikre, når mindre grupper betragtes Der hvor den store variation i vægtene giver problemer er, når man ønsker at betragte en mindre gruppe, som f.eks. personer uden arbejde eller en specifik aldersgruppe. Når man betragter en subpopulation af stikprøven på 5.500 husstande, så er gruppen/stikprøven, som betragtes, naturligvis mindre. Det resulterer i sig selv i større stikprøveusikkerhed. Kalibreringen sikrer, at nogle få

subpopulationer holdes på plads, som fx risiko for fattigdom på aldersgrupper. Betragtes andre kryds i stikprøven, øger variation i vægtene stikprøveusikkerheden væsentligt. Skal mere detaljerede indikatorer ovenikøbet kunne bruges på regioner (NUTS 2 niveau), er det meget vigtigt at få begrænset variationen på vægtene mest muligt.

Det første af revisions elementer til begrænsning af vores stikprøveusikkerhed er:

- Skift af familiebegreb i kalibreringen.

SILC er siden 2009 blevet vægtet op imod Danmarks Statistiks såkaldte e-familier og før det på adresser. Familiebegrebet skifter altså fra det relativt snævre e-familiebegreb, til et nyt bredere familiebegreb.

Et nyt familiebegreb

Det nye familiebegreb er dannet med udgangspunkt i e-familien. Her får fx hjemmeboende over 25 år deres egen familie, selvom de bor på samme adresse som deres forældre. Det nye familiebegreb sikrer, at alle på adressen grupperes i samme familie, hvis der er blot en enkelt biologisk forbindelse i to generationer mellem to e-familier. Det vil sige at bedsteforældre, onkler, nevøer og deres e-familie på samme adresse alle antages, at tilhøre samme familie på adressen.

Det næste af revisions elementer til begrænsning af vores stikprøveusikkerhed er:

- Ændring af stikprøveudtrækning (fra SILC 2015)

Ny metode til udtrækning af stikprøven

Herudover er gennemført en ændring af stikprøveudtrækning. Fra 2015 er stikprøveudtrækningen ændret, så der nu først trækkes en adresse og derefter en person på adressen. Tidligere er personen udtrukket direkte. Tidligere praksis har medført, at store familier har været overrepræsenteret i stikprøven, idet deres udtrækningssandsynlighed var større end fx enliges. Da SILC først og fremmest vægtes på husstands-niveau, vil denne ændring, når den er fuld indfaset i 2018, yde det væsentligste bidrag til at reducere spredningen på vægtene.

1.2 Sammenlignelighed med øvrige lande

Danmark er det land med suverænt flest negative indkomster i SILC. Det skyldes, at Danmark muligvis som det eneste land har sikret, at ekstreme indkomster, positive såvel som negative, repræsenteres fuldt i SILC.

At medtage negative indkomster fuldt, giver på den ene side det mest korrekte billede af danske indkomstforhold. På den anden side giver det problemer med sammenligneligheden, idet flere lande ikke kan rapportere negative indkomster. Det kan være på grund af manglede skatte-registrering af negative indkomster eller blot fordi, at landene ikke anerkender negative indkomster og derfor sætter dem til nul. I andre lande er SILC udelukkende stikprøvebaseret. I sådanne tilfælde er det almindelig praksis at benytte en form for top/bottom-coding, hvor der ligges et arbitrært loft over ekstreme indkomster for at mindske usikkerheden på data. Praksis varierer meget fra land til land.

For at komme mere på linje med de øvrige lande indføres derfor:

- Loft over negativ ækvivaleret indkomst i kalibreringen på €100.000 kr. (745.000 kr. i 2013-priser)

Loftet indføres kun i kalibreringen. Der kan altså fortsat indgå personer med en negativ indkomst over €100.000 i SILC. Men personerne med meget høje negative indkomster vil få en lavere vægt. Det nedbringer den samlede mængde af negative indkomster (se figur 1.1 i næste afsnit).

Et andet punkt, hvor Danmark skiller sig lidt ud fra de øvrige lande, er at vi har kalibreret stikrøven på et relativt snævert familiebegreb, hvor fx hjemmeboende over 25 år får tildelt deres egen familie (e-familier). Derfor er det vurderingen, at skiftet af familiebegreb også bidrager til forbedring af sammenligneligheden i SILC (se afsnit 1.1).

1.3 Indkomst ændringer

Som opfølgning på sidste års revisioner af det nationale indkomstregister og revisionen af det danske nationalregnskab ændres følgende:

- Kirkeskat betragtes ikke længere som en skat, men som et frivilligt bidrag. Den indgår derfor ikke længere i skattebetalingen og trækkes heller ikke fra den disponible indkomst.
- Grøn check og varmemhjælp inkluderes i indkomsten.
- Niveauet for lejeværdi af egen bolig reduceres som følge af en tilsvarende niveauændring i nationalregnskabet.

Kun de to første komponenter vil påvirke den disponible indkomst i SILC, idet lejeværdi af egen bolig ikke er en del af SILC's disponible indkomst.

1.4 Øvrigt

- Familier, hvor alle indkomstkomponeanter er inden for intervallet [-2.500; 2.500] kr., fjernes fra registret under vægtningen.
- Vægtene trimmes. Det vil sige, at det bånd, hvor de mindste og største vægte skal ligge inden for, indsnævres lidt.

Herudover vil der, i det omfang, vi er blevet gjort opmærksomme på dem, blive rettet op på kendte fejl og mangler på nogle af SILC's mere end 200 variable, efterhånden som SILC føres tilbage i tid. Det drejer sig blandt om variabelen PLO31 i SILC's longitudinelle del i perioden 2006-2010, samt korrektion af tilhørsforhold til husstanden for en mindre gruppe personer i 2013.

2. Uddybet formål og effekt af familieændringer og loft over negative indkomster

Adresser frem for
E-familier

Vægtningen er hidtil foretaget op imod Danmarks Statistiks e-familier, som generelt benyttes i DST. E-familien er et meget snævert familiebegreb, da fx hjemmeboende over 25 år får deres egen familie. Målet med revisionen af familiebegrebet har været at finde det bredest mulige familiebegreb, uden at gå hele linen ud og benytte adressen. Ved brug af adressen ville alle inkl. logerende, au-pair mv. antages at tilhøre samme husstand og have del i husstandens velstand eller mangel på samme.

Derfor udvides e-familien så alle, som har en biologisk forbindelse på samme adresse op til en fælles bedsteforældre, knyttes sammen i en familie. Det vil sige, at alle hel- og halvsøskende, samt onkler, tanter, nevøer antages at tilhøre

samme husstand, hvis de bor på samme adresse ifølge befolkningsregistret. Den nye familietype benævnes *SILC-familier* fremadrettet.

Tabel 1 **Familiebegreber og match med selvrapporterede SILC-husstande**

	Familier i befolkningen	Gns. medlemmer pr. familie	Match med SILC husstande
	antal		pct.
E-familier	2 924	1,927	89,82
SILC-familier	2 840	1,982	93,86
Adresse_id	2 625	2,147	95,10

I SILC-interviewet definerer respondenterne selv sin husstand. Som det fremgår af tabel 1, er det adressen, som matcher de selvrapporterede SILC-husstande bedst. Tabellen viser, at de personer, som respondenterne har angivet tilhører hans/hendes husstand matcher de personer, som ifølge registret bor på adressen i 95 pct. af tilfældene. De nye SILC familier matcher for næsten 94 pct. af respondenterne mens de e-familier, som hidtil er brugt i kalibreringen kun matcher for 90 pct. af respondenterne.

Det skal dog i denne forbindelse medtages, at respondenterne i interviewet bliver mødt med et forslag til husstandens medlemmer, som netop er dem, der bor på adressen. Det giver måske et lille bias til fordel for match på adressen, idet respondenterne i tvivlstilfælde, måske tager det præsenterede forslag for pålydende.

Når SILC-familierne og ikke adressen er valgt, selvom den matcher en smule bedre end SILC-familierne, skyldes det til dels mistanken om det førnævnte bias, men også et ønske om at sikre, at loggerende, adresser med flere familier på, au-pair og andre tilfælde, hvor der oftest er tale om to adskilte husstande i økonomisk forstand på adressen også indregnes i SILC som separate husstande.

Internationalt er udgangspunktet for kalibreringen oftest enten adressen eller respondent-definerede husstande fra nationale folketællinger. Hvis de respondent-definerede husholdninger i folketællingerne følger mønstret, når respondenterne definerer deres husstand i den danske SILC, så må disse familiedefinitioner også formodes at have væsentlig bredere familiebegreber end de danske e-familier.

Effekt Da det nye familiebegreb ligner de selvdefinerede SILC-husstande væsentligt mere end e-familien, bidrager dette også til at begrænse spredningen på vægtene i SILC.

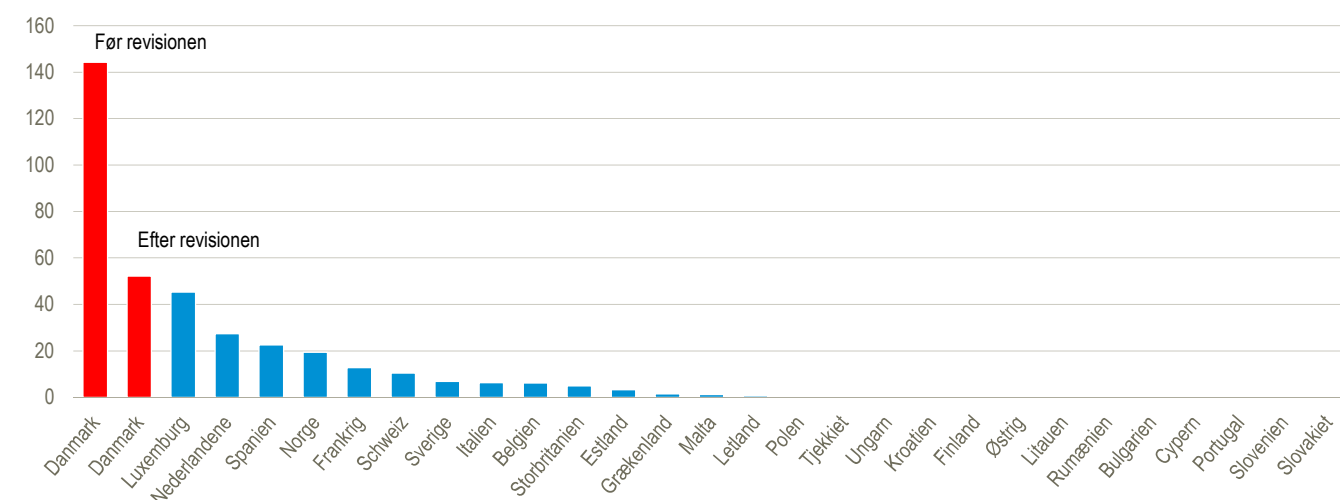
Vægtning på de nye SILC-familier frem for e-familien vil betyde færre familier i SILC (vægtet). Dette bidrager i retning af en lille stigning i den gennemsnitlige familieindkomst. De større familieenheder i kalibreringen bidrager også til en lille mindskning af den målte indkomstulighed.

Loft over negative indkomster i registret

Personer med ekstreme indkomster er oftest underrepræsenteret i undersøgelser som SILC. Som et af de eneste lande, vægtes der i den danske udgave på indkomstmassen, hvilket betyder, at negative indkomster i registret er fuldt repræsenteret i den danske SILC.

Praksis varierer	Flere lande benytter top/bottom-coding, hvor man ligger et loft over og under på ekstreme indkomster, for ikke at få for stor stikprøveusikkerhed. Praksis er her meget forskellig fra land til land. Herudover kan der være forskelle på skattesystemer, hvor fx negativ indkomst i nogle lande blot sættes til nul af skattevæsenet og enten ikke kommer ind i systemet eller først registreres som fradrag i efterfølgende år. Endeligt er det også muligt, at nogle lande blot ikke anerkender og opererer med negative indkomster. I Danmark registreres alle negative indkomster, og de registreres vel og mærke i det år, indkomsten er tabt. Endelig kalibreres i Danmark ikke kun på antallet med negativ/lav indkomst, som i flertallet af lande med kalibrering på indkomsten, der kalibreres også på indkomstmassen. Det har sikret, at også ekstreme indkomster har været talt fuldt med. Den danske praksis har sikret, at omfanget af negative indkomster er målt helt korrekt. Men konsekvensen er dårligere sammenlignelighed med de øvrige lande i SILC, idet Danmark står alene med denne praksis og år efter år har været det land med suverænt flest negative indkomster. Dette er et problem for sammenligneligheden, da negative indkomster fx påvirker ulighedsmål, som baserer sig på indkomstmassen. Det er fx gini-koefficienten og i særlig grad 80/20-raten. Derfor er et loft på ækvivaleret disponibel indkomst på omkring €100.000 (745.000 kr.) introduceret i vægtingen af SILC.
Effekt af loft over negative indkomster	Som figur 1 viser, er Danmark med loftet over negative indkomster fortsat det land med flest negative indkomster i SILC 2013, men skiller sig ikke længere så kraftigt ud fra de øvrige lande. Figur 1 viser effekten på gennemsnittet for ækvivaleret disponibel indkomst, hvis negativ indkomst for alle lande blev sat til nul (som det fx er praksis i OECD's opgørelser). Målt på det fulde register, vil det, at sætte negative indkomster til nul, bidrage til at forøge gennemsnitsindkomsten i Danmark med €142 i 2013. Dette er væsentlig mere end i noget andet land i SILC, hvor mange lande endda ikke opererer med negative indkomster. Med loftet over negative indkomster vil reduktionen af gennemsnitsindkomsten kun være på €52 ved samme øvelse. Danmark vil hermed fortsat være blandt de lande med flest negative indkomster i 2013, men skiller sig mindre ud fra de øvrige lande med det arbitrært fastsatte loft over negative indkomster på €100.000.
Effekt af loft over negative indkomster	Den isolerede effekt af et loft over negative indkomster på €100.000 er for indkomståret 2012 (SILC 2013), at den gennemsnitlige ækvivalerede disponible indkomst måles €92 (690 kr.) højere end tilfældet ville have været uden loftet. Det påvirker ligeledes uligheden, fx bidrager loftet isoleret til et fald i gini-koefficienten på 0,2 gini-point i 2013.

Figur 1 Effekten af at sætte negative indkomster til nul på den gennemsnitlige ækvivalerede indkomst for personer. SILC 2013. Før og efter revisionen.



Kilde: Egne beregninger på baggrund af SILC User database 2013 og indkomstregistret.

Hertil skal dog siges, at effekten af loftet vil være væsentlig større omkring finanskrisen. Her led ekstraordinært mange personer meget store tab og omfanget af negative indkomster var ca. fire gange større målt på disponibel indkomst i 2008 i forhold til 2013 i Danmark.

Resultater før og efter revisionen

Strukturvægtningen er gennemført før og efter revisionen. I dette afsnit vises effekten af revisionen på vægtenes variation og på udvalgte indikatorer.

Negative vægte accepteres ikke

Til vægtningen benyttes SAS programpakken Clan. Husstandstyper, som er underrepræsenteret i stikprøven, får en høj vægt, og omvendt får de overrepræsenterede husholdninger en lav vægt. Vægtningen sikrer, at antallet af personer i foruddefinerede grupper i stikprøven matcher befolkningen når de multipliceres med deres vægte, samtidig med at variationen på vægtene reduceres.

Da kalibreringen er relativ kompliceret og stikprøven, grundet bortfald og metode til sampling, har været lidt for skæv i forhold til befolkningen, trækkes der så meget i vægtene, at de bliver negative. Negative vægte accepteres ikke, de sættes til et minimumsniveau og fastholdes, og vægtningen køres igen på den resterende del af befolkningen.

Spredning på vægte

Både antallet af negative vægte, som Clan skal rette op og spredningen i vægten bliver væsentlig mindre. Revisionen betyder isoleret, at vægtningsprogrammerne kun bruger fire iterationer mod tidligere fem og antallet af vægte falder fra 251 til kun 57 efter revisionen. Det på trods af at kravet til minimums niveau på vægtene er hævet.

Tabel 3 viser, at spændet mellem de laveste og højeste vægte er reduceret ganske betragteligt. Spredningen på vægtene vil blive yderligere reduceret de kommende år, som følge af ændringerne i stikprøve udtrækningen.

Tabel 3 Spredning på vægte. 2014

1st Pctl	5th Pctl	10th Pctl	50th Pctl	90th Pctl	95th Pctl	99th Pctl	Std Dev
----------	----------	-----------	-----------	-----------	-----------	-----------	---------

Før revisionen	2	3	62	286	769	1 056	1 801	416
Efter revisionen	40	86	128	312	744	955	1 747	357

Revisionen vil blive ført tilbage til 2008 i løbet af det kommende år.

Revisionens samlede effekt på de væsentligste SILC-indikatorer vil blive tilføjet denne revisionsbeskrivelse, når revisionen er gennemført for 2013.