

Om usikkerhed i EU-SILCs målinger af indkomstfordelingen

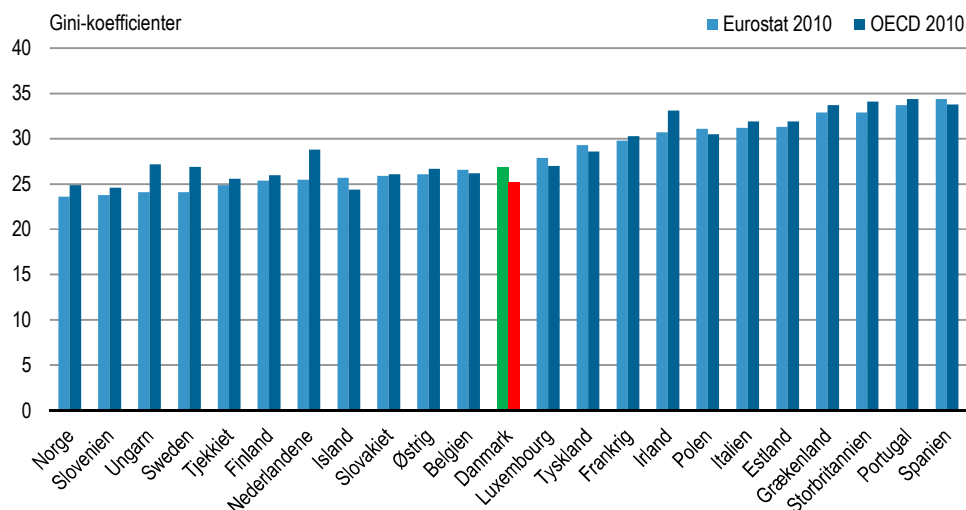
Til brugere af indkomststatistikken

Januar 2017: Siden dette notat er skrevet er der gennemført en revision af SILC og sket ændringer i lovmodellen. Derfor kan enkelte beskrivelser i notatet være uaktuelle. De overordnede konklusioner holder dog fortsat. Når revisionen er ført tilbage i tid vil dette notat blive erstattet af et nyt. Forventeligt i løbet af efteråret 2017.

Samme mål,
forskellige resultater

Internationale sammenligninger af fx. Gini-koefficienten giver ofte forskellige resultater afhængig af, hvilken kilde og dataår der betragtes. Dette skyldes forekomsten af databrud, varierende opgørelsesmetoder og betydelig statistisk usikkerhed på nogle af de mest anvendte fordelingsmål.

Gini-koefficienter målt af Eurostat og OECD



Anm.: For Irland og Ungarn er OECD tal for 2009.

Dette notat belyser nogle af kilderne til den statistiske usikkerhed, ligesom det beskriver et databrud i Eurostats serier for Danmark. Endelig beskrives sammenhængen mellem Eurostats, de nationale og OECD's opgørelser af Gini-koefficienten.

Om levevilkårsundersøgelsen

Introduktion af
levevilkårsundersøgelsen

SILC står for *Statistics on Income and Living Conditions* og betegnes på dansk som levevilkårsundersøgelsen. Det er en stikprøveundersøgelse, som årligt gennemføres i alle EU-lande og en række andre europæiske lande udenfor EU. Levevilkårsundersøgelsen gennemføres af de nationale statistikbureauer. I Danmark af Danmarks Statistik. Danmark har deltaget siden starten i 2003. I Danmark deltager hvert år i alt omkring 5.500 husstande.

Levevilkårsundersøgelsen indeholder omkring 200 variable, der med forskellig detaljeringsgrad beskriver forhold om:

- indkomst
- materielle afsavn
- status på arbejdsmarkedet
- børnepasning
- uddannelse
- boligforhold
- helbred
- subjektiv vurdering af husstandens økonomi mv.
- et årligt modul med varierende emner.

Indsamlingsmetode	Langt de fleste af disse 200 variable, som skal indsamles for undersøgelsens deltagere, hentes fra Danmarks Statistiks registre, men ikke alle variable kan udfyldes ved brug af registre. Der gennemføres derfor et interview på web, telefon eller med papirskema, hvor deltagerne spørges om de emner, Danmarks Statistik ikke kan belyse ved hjælp af registre.
Non-response	Deltagelse i undersøgelsen er frivillig. En betydelig andel af de personer, som Danmarks Statistik udtrækker, vælger ikke at medvirke i undersøgelsen (<i>non-response</i>). Normalt deltager kun lidt over 50 pct. af de udtrukne personer i undersøgelsen.
Vægtning af stikprøven	<p>For at minimere den statistiske usikkerhed og korrigere for et skævt non-response tildeles deltagerne i levevilkårsundersøgelsen en vægt. Denne vægt sikrer, at stikprøven på en række parametre ligner befolkningen mest muligt.</p> <p>Disse parametre er:</p> <ul style="list-style-type: none">• alder• køn• husstandstype• socioøkonomisk status• uddannelse• indkomstintervaller - herunder Risiko for fattigdom (antal under 60 pct. af medianindkomsten)• indkomstmassen indenfor indkomstintervaller (fra 2010)
Lille usikkerhed på variable, som indgår i vejningen	Efter vægtning er stikprøveusikkerheden på disse variable, isoleret set, stort set ikke eksisterende. Kombinerer man derimod flere parametre fx risiko for fattigdom med alder, risikerer man, at en lille usikkerhed opstår for de enkelte aldersgrupper, mens usikkerheden for det samlede antal i risiko for fattigdom er minimal.
Struktur og forløb	<p>Levevilkårsundersøgelsen er inddelt i en struktur- og en forløbsdel. Levevilkårsundersøgelsen er en panelundersøgelse, hvor deltagerne bliver bedt om at deltage fire år i træk, med løbende udskiftning af panelet. De af deltagerne, som har deltaget flere år i træk, udgør forløbsdelen af levevilkårsundersøgelsen. Strukturdelen er et tværsnit, hvor alle deltagerne indgår uanset om de har deltaget et eller flere år.</p> <p>Mens der er omkring 5.500 deltagende husstande i strukturdelen, består forløbsdelen ofte kun af 600-800 husstande i Danmark.</p>

Forløbsvægtene er i vid udstrækning baseret på strukturvægtene korrigeret for ændringerne i stikprøvens sammensætning afhængig af, om vi betragter gruppen, som har deltaget de seneste to år, tre år eller fire år.

Flertal af indikatorer beregnes på strukturdelen

Flertallet af de SILC-indikatorer, som findes i [Eurostats database](#), er baseret på undersøgelsens strukturdelen. Dette gælder i forhold til indkomstfordelingen bl.a. for Gini-koefficienten, risiko for fattigdom og 80/20-raten. Indikatorer som varig risiko for fattigdom (*persistent Risk of Poverty*), hvor man ser på andelen af personer, som har haft under 60 pct. af medianindkomsten i indeværende år og mindst to af de tre foregående år, er baseret på forløbsdelen.

Statistisk usikkerhed og databrud

Statistisk usikkerhed

Da levevilkårsundersøgelsen er en stikprøveundersøgelse, er den på trods af vægtningen forbundet med en vis statistisk usikkerhed. Størrelsen af denne usikkerhed afhænger både af, hvilken indikator og undergruppe man ser på.

... på ofte anvendte indikatorer

Risiko for fattigdom (ROP) er en indikator med meget lille usikkerhed, idet den indgår direkte i stikprøvevejningen. Gini og 80/20 - raten er derimod eksempler på indikatorer, som før 2010 har afvejet fra de nationale tal pga. almindelig statistisk usikkerhed og et bias (systematisk fejlmåling). Det bias, som er opstået, er primært et resultat af utilstrækkelig korrektion for de lavere deltaerprocenter for personer med meget høje og lave indkomster. Dette er resultatet i at fx Gini-koefficienten og 80/20-raten er målt lidt for lavt.

Afviselser mellem offentliggjort stikprøves og de faktiske Gini-koefficienter

Tabellen nedenfor sammenligner Gini-koefficienten i Eurostats database med den faktiske Gini-koefficient målt for hele befolkningen med anvendelse af ens metoder og indkomstbegreber. Som det fremgår, varierer afstanden mellem Gini-koefficienten i EU-stikprøven og registret før 2010 betydeligt, med en Gini-koefficient, som i Eurostats database er mellem 1 og 1,7 Gini-point for lav.

Tabel 1.

Afviselser mellem offentliggjorte og faktiske Gini-koefficienter

SILC-år	2007	2008	2009	2010	2011	2012
Registerår	2006	2007	2008	2009	2010	2011
EU-SILC stikprøve	25,2	25,1	26,9	26,9	27,8	28,1
Gini - hele befolkningen (SILC metoder og indkomstbegreb)	26,7	26,8	27,9	26,9	27,8	28,0
Afvigelse mellem faktisk Gini og stikprøve Gini	1,5	1,7	1,0	0,0	0,0	0,1

Ny vægtningsmetode og databrud i 2010

Som konsekvens af dette introducerede Danmarks Statistik i Danmark i 2010-undersøgelsen indkomstmassen som en del af vægtningen i levevilkårsundersøgelsen. Indkomstmassen er summen af indkomstbeløbene indenfor en række af indkomstintervaller. Da det netop er fordelingen af indkomstmassen, som Gini-koefficienten beskriver, betyder det, at afvigelsen mellem den faktiske Gini-koefficient målt på hele befolkningen og Gini-koefficienten, som offentliggøres af Eurostat, er væsentligt formindsket.

Stigning fra 2007-2009 til 2010-2012 blev opgjort for stor

Dette medfører også at stigningen i Gini-koefficienten fra perioden før 2010 til perioden fra 2010 og frem ser lidt større ud, end den reelt var. Altså bør man som bruger af Eurostat-data være opmærksom på de svingende Gini-koefficienter i perioden op til 2010 og på databrudet i 2010, som gør det svært at tolke præcis, hvor meget Gini-koefficienten er steget fra årene før 2010 og til de efterfølgende år.

Gini-koefficient og 80/20-rate rammes særligt af databrud i 2010

Denne usikkerhed før 2010 og databruddet fra 2009 til 2010 påvirker de fordelingsmål, hvori indkomstmassen indgår. Dette er særligt Gini-koefficienten og 80/20-raten. Sammenligninger af indkomsten mellem forskellige grupper ved brug af gennemsnitsindkomsten (*mean income*) påvirkes også i begrænset omfang af dette databrud.

Andre fordelingsmål, som fx. risiko for fattigdom og medianindkomster, er til gengæld stort set upåvirket af føromtalt ændring.

Tabel 2. Effekt af usikkerhed og databrud på de enkelte indikatorer

Indikator	Baseret på	Effekt af databrud 2010	Vurderet omfang af statistisk usikkerhed
Gini-koefficient Gini coefficient	Struktur	Betydelig	Før 2010: betydelig Fra 2010: minimal
80-20 raten S80/S20	Struktur	Betydelig	Før 2010: betydelig Fra 2010: minimal
Risiko for fattigdom Risk of Poverty(ROP)	Struktur	Minimal	Minimal
Varig risiko for fattigdom Persistent Risk of Poverty	Forløb	Minimal	Betydelig
Medianindkomst Median income	Struktur	Minimal	Minimal
Gennemsnitsindkomst Mean income	Struktur	Lille	Før 2010: Lille Fra 2010: Minimal
Afstand til grænse for risiko for fattigdom (for personer i risiko for fattigdom) Poverty gap	Struktur	Minimal	Betydelig

Underinddelinger af befolkningen giver større usikkerhed

Den anden kilde til usikkerhed opstår, hvis man betragter små undergrupper af EU-SILC. Des mindre gruppen, som betragtes er, jo større er den statistiske usikkerhed. Dette er blandt andet tydeligt, når man betragter *Persistent Risk of Poverty*, der som nævnt kun er baseret på 600-800 husstande. Særligt når man fx betragter specifikke aldersgrupper, vokser usikkerheden betydeligt, da dette yderligere begrænser stikprøvens størrelse, og vægtningen på så detaljeret niveau ikke kan holde indikatoren på plads.

Et andet eksempel på dette er det såkaldte *Poverty gap*. Her måles afstanden fra medianen for personer under fattigdomsgrænsen, op til grænsen ved 60 pct. af medianindkomsten. Denne beregnes i sagens natur kun for de omkring 13 pct. af de 5.500 deltagende husstande, som befinder sig under denne grænse. Deles denne yderligere op på fx alder og køn, bliver denne stikprøve meget lille og dermed præget af stor statistisk usikkerhed.

Forskelle mellem den nationale og SILCs Gini-koefficient

Danmarks Statistiks officielle Gini-koefficient

Danmarks Statistik beregner Gini-koefficienten baseret på data for hele den danske befolkning. Denne tal-serie kan blandt andet findes i Statistisk Tiårs-oversigt og vores tema publikation [Indkomster 2011](#).

Opgørelsesmetoderne er stort set ens. Der er dog fire væsentlige forskelle på de to opgørelsesformer:

Tabel 3. Kilder til afvigelser mellem SILC og den nationale opgørelse

	EU SILC	National opgørelse	Konsekvens
Population	Stikprøvebaseret	Fuld population	Statistisk usikkerhed (stort set elimineret fra 2010 og frem)

Indkomster er fra år	år-1	år	Dvs. at SILC-tal fra 2010 skal sammenlignes med de danske 2009-tal, som dækker over indkomståret 2009.
Indkomstbegreb	Inkluderer ikke lejeværdi af egen bolig og prioritetsrenter i deres indkomstbegreb	Lejeværdi og prioritetsrenter inkluderet	Betyder, at uligheden vurderes højere i EU-SILC end nationalt.
Familiebegreb	Benytter husstande, som består af personer bosat på samme adresse med fælles økonomi	Benytter Danmarks Statistiks E-familier	Husstande er oftest en større enhed end Danmarks Statistiks E-familier. Effekten på den målte ulighed er begrænset.

I 2012 (indkomståret 2011) er Gini-koefficienten for Danmark i Eurostats opgørelse på 28,1. Den nationale opgørelse af Gini-koefficienten for 2011 er 28,0.

Overensstemmelse mellem officiel Gini og SILC Gini siden 2010

Overordnet har der siden 2010 været fin overensstemmelse mellem den Gini-koefficient, som er beregnet i levevilkårsundersøgelsen, og den nationale opgørelse, da ovenstående effekter nogenlunde har udlijnet hinanden. Men der vil kunne forekomme små afvigelser mellem de to opgørelser som følge af de ovenfor nævnte forskelle.

Forskelle mellem SILC og OECD-tal

Lovmodellen og OECD

Som det ses på figuren på første side, er der heller ikke altid fuld overensstemmelse mellem Eurostat og OECD tal. Det er Økonomi- og Indenrigsministeriets *lovmodel* (tidligere finansministeriet), som står bag de danske OECD-tal.

Lovmodellens disponible indkomst og Danmarks Statistiks definition af den disponible indkomst er ikke helt ens. Forskellene drejer sig bl.a. om behandlingen af private pensionsopsparinger, indkomsten for de selvstændige. ~~og det forhold at lovmodellen ikke medtager personer med de mest ekstreme høje og negative indkomster.~~

Samlet bevirker disse forskelle, at Danmarks Statistik oftest måler en lidt højere ulighed end *lovmodellen*.

Derudover er der følgende forskelle mellem SILC's og OECD's opgørelser:

Tabel 4. Sammenligning af metoder mellem OECD og Eurostat

	Eurostat(SILC)	OECD
Indkomstbegreb	Danmarks Statistiks disponible indkomst (uden lejeværdi og prioritetsrenter)	Lovmodellens disponible indkomst (uden lejeværdi og alle renteudgifter)
Ækvivalens metode	OECD's modificerede skala: Første voksen: 1 Øvrige voksne (over 14 år): 0,5 Børn (under 15 år): 0,3	(Antal personer i husstanden) ^{0,5}
Datakilder	EU-SILC (Survey) Omkring 5.500 husstande	Blandede datakilder, for de fleste lande EU-SILC eller forbrugsundersøgelsen DK: Lovmodellen er også en stikprøve (33 pct. af befolkningen). Men stikprøven er væsentligt større end EU-SILC og har intet bortfald, da alle oplysninger er tilgængelige i registret.
Dataår	DK: År -1	DK: år+0
Fejkilder / usikkerhed	Se ovenfor	OECD opererer, i modsætning til eurostat ikke med negative indkomster. OECD's opgørelse er for de fleste

Anbefalinger

- Betydelig usikkerhed og risiko for bias i øvrige lande Da Danmark er et af de få lande, som i det hele taget har et indkomstregister at vægte stikprøven op imod, må det antages, at usikkerheden og risiko for bias på fordelingsmålene i de øvrige EU-lande er større end usikkerheden på de danske tal har været.
- Forskellige indkomstbegreber I tillæg til dette er der trods de bedste bestræbelser på at lave sammenlignelige indkomstbegreber mellem landene også betydelige forskelle her. Dette komplicerer yderligere de internationale sammenligninger.
- Brug tal med forsigtighed og tag nødvendige forbehold Danmarks Statistik anbefaler derfor, at små forskelle i de internationale sammenligninger af indkomstfordelingen tolkes med forsigtighed. Dette gælder både i forhold til rangering af landene efter (u)lighed i de enkelte år, men også når udvikling over tid sammenlignes på tværs af landegrænser.
- Anvend nationale serier, hvor det er muligt Danmarks Statistik offentliggør blandt andet i *Statistisk Tiårsoversigt* og publikationen *Indkomster* en række fordelingsmål. Det anvendte indkomstbegreb er ikke præcis det samme, som benyttes af Eurostat, og derfor bør disse ikke benyttes til internationale sammenligninger. Men Danmarks Statistik anbefaler, at man i videst muligt omfang benytter sig af disse nationale indkomstopgørelser, når udviklingen i indkomstuligheden opgøres.
- Pas på usikkerhed, når der måles på små befolkningsgrupper De stikprøvebaserede indikatorer i Eurostats database er behæftet med betydelig usikkerhed, som bliver større, des mindre de delpopulationer der betragtes er. Derfor skal man passe på med at overfortolke små forskelle på udvikling såvel som niveau for en del af disse indikatorer.
- Mere information Læs mere om de nationale fordelingsmål i publikationen *Indkomster* på www.dst.dk/publ/indkomster
- Se Eurostats database på:
http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database