

Det standardberegnete lønindeks. Metode efter serviceeftersyn 2022-2023

1 Rammen for serviceeftersynet

Danmarks Statistik har i 2022-2023 gennemført et serviceeftersyn af det standardberegnete lønindeks. Det primære fokus for serviceeftersynet var at gennemgå beregningsprogrammerne, der anvendes til det standardberegnete lønindeks. De indledende databehandlingsprocesser fra indberetningerne modtages af Danmarks Statistik, frem til data lægges over i denne del af produktionssystemet var ikke omfattet af serviceeftersynet. I forbindelse med serviceeftersynet blev der imidlertid konstateret flere uhensigtsmæssigheder med de indledende databehandlingsprocesser. Systemerne er relativt gamle, utilstrækkeligt dokumenterede og ikke designet med et helhedsperspektiv i forhold til at sikre datakvaliteten og understøtte den videre brug af data. Dette forhold er ikke afgrænset til det standardberegnete lønindeks men gældende for alle lønstatistikkerne (lønindeks og lønstrukturstatistikken).

Der er en række andre elementer, der ligeledes har været et grundvilkår for arbejdet med serviceeftersynet af det standardberegnete lønindeks. Dette gælder fx stikprøvedesignet, den aktuelle responsrate blandt indberetterne og det generelle formål med indekset. Indeksets formål blev formuleret mere præcist ifm. serviceeftersynet, men formålet er grundlæggende uændret før og efter serviceeftersynet (se nedenfor).

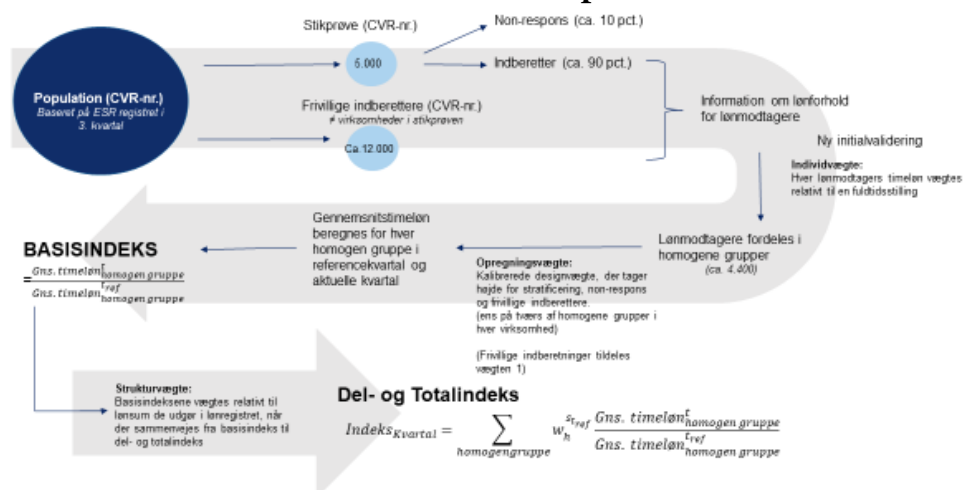
Centrale elementer og forudsætninger for serviceeftersynet af det standardberegnete lønindeks er vist i figur 1. Den resterende del af nærværende notat følger den rækkefølge, der er angivet i figur 1.

Figur 1. Rammer for serviceeftersynet af det standardberegnete lønindeks



I figur 2 vises en samlet model for produktionssystemet der danner det standardberegnete lønindeks. Man kan her se, hvordan der fra populationen af private enheder trækkes en stikprøve, der til slut opregnes til del- og totalindeks.

Figur 2: Den fulde model – overblik for den private del



2 Præcisering af formål

I forbindelse med serviceeftersynet af det standardberegnete lønindeks blev formålet med lønindekset præciseret:

Formålet med det standardberegnete lønindeks er at belyse lønudviklingen for lønmodtagere i Danmark, så vidt muligt rensat for ændringer i beskæftigelsessammensætningen på arbejdsmarkedet.

Kilde: Statistikdokumentationen til det standardberegnete lønindeks

Tidligere skrev Danmarks Statistik i formålet, at der blev korigeret for strukturelle ændringer. Dette var ikke specificeret, men forståelsen var, at der skulle korrigeres for forskydninger i beskæftigelsessammensætningen og for forskydninger mellem de indberettende virksomheder.

Forståelsen af korrektion for forskydninger i beskæftigelsessammensætningen og mellem virksomheder har fejlagtigt været, at korrektionen skete gennem et match på cvrnr. og arbejdssted, så kun de virksomheder, som havde indberettet to kvartaler i træk var med i kvartalsstigningen.

Reelt sker korrektionen for forskydningerne via homogene grupper og tilhørende basisindeks.

3 Stikprøve og designvægte

Stikprøvedesignet er som tidligere nævnt en forudsætning for serviceeftersynet. Stikprøven til lønindeks på ca. 5.000 virksomheder er trukket blandt de virksomheder, som er forpligtigede til at indberette til lønstrukturstatistikken (alle virksomheder med mere end 10 ansatte).

3.1 Stikprøvedesign

Stikprøven trækkes i strata efter branche og størrelse. Lønkomponenter og personoplysninger fra lønmodtagere i de udtrukne virksomheder indberettes.

De udtrukne virksomheder indberetter data for alle deres medarbejdere. Der udvælges således ikke "varer" i form af særlige medarbejdergrupper på de enkelte virksomheder (homogene grupper), hvor det kun er disse foruddefinerede medarbejdergrupper der skal indberettes for.

Dette er metodemæssigt relevant og væsentligt i forhold til en sammenligning med fx forbrugerprisindekset, hvor der netop udtrækkes en varekurv. Implikation er nemlig, at det ikke er muligt at anvende imputering ift. non-respons, idet det ved beregningen er uvidt, om data om en given homogen gruppe, der ikke er i materialet for et kvartal men var der i det foregående kvartal, skyldes (1) en ændring i beskæftigelsessammensætningen eller (2) er en følge af manglende indberetning.

3.2 Opregning

I det standardberegnete lønindeks korrigeres for manglende indberetninger via opregningsvægte. Virksomheder, der har indberettet, får korrigeret (kalibreret) deres oprindelige vægt (designvægten) for at kompensere for, at andre lignende virksomheder i det pågældende stratum ikke har indberettet.

Korrektion for manglende svar gennem vægtning er et normalt metodisk standardvalg. Metoden er et alternativ til imputering, idet de manglende værdier i stedet indirekte erstattes med gennemsnitstimelønnen for de øvrige lønmodtagere i samme homogene

gruppe gennem vægtkalibreringen. Man kan således ikke direkte se eller beregne de imputerede værdier for de manglende virksomheder.

For hvert kvartal fastholdes opregningen for de tidligere kvartaler, så der sker ikke ændringer i de beregnede indeksværdier bagud i tid – på trods af, at der efter indberetningsfristen kan være indberettet yderligere.

3.3 Frivillige indberettere

Ud over de udtrukne virksomheder modtager Danmarks Statistik også frivillige indberetninger. Ved trækningen af stikprøven er der ikke taget højde for, om en virksomhed i forvejen indberetter frivilligt. Dette betyder, at en frivillig indberetter, som også er udtrukket, vægtes til at repræsentere andre virksomheder, som ligner den pågældende virksomhed. Øvrige frivillige indberetninger fra virksomheder, som ikke er trukket til stikprøven, indgår stadig i beregningen, men repræsenterer kun sig selv.

Begrundelsen for dette er, at de frivillige indberettere udgør en særlig gruppe, der bl.a. er karakteriseret ved, at de anvender en specifik løndataleverandør, eller at de frivilligt leverer deres løndata til Dansk Arbejdsgiverforening. Det kan derfor ikke udelukkes, at de adskiller sig fra andre virksomheder mht. løn eller fraværsmønster. Stikprøver, som ikke er udvalgt tilfældigt, er ikke underlagt almindelig sandsynlighedsbaseret stikprøveteori, og der kan derfor ikke vægtes op og beregnes usikkerhed som normalt.

Omfanget af frivillige indberetninger er på ca. 12.000 virksomheder, som skal lægges til de ca. 5.000 fra stikprøven.

3.4 Offentlige løndataleverandører

I tillæg til stikprøven af virksomheder, der trækkes fra private løndataleverandører, modtager Danmarks Statistik information fra offentlige løndatasystemer. I nogle tilfælde indgår enheder i en anden økonomisk sektor i de offentliggjorte tal i forhold til den løndataleverandør, der er kilde til indberetningen.

I forhold til den private sektor i form af virksomheder og organisationer vil data, der kommer fra offentlige løndataleverandører, der indgår i beregningen for den private sektor, således blive betragtet som frivillige indberettere.

4 Responsrate og bortfald

Det er et faktum, at der er et non-responsrelateret bortfald i forbindelse med indsamlingen af data. Bortfaldet udgør i størrelsesordenen 10 pct. af virksomhederne hvert kvartal. Målt på lønmodtagere svarer det bortfald, der kan relateres til non-respons i stikprøvevirksomhederne, gennemgående til 5-6 pct. af det samlede antal lønmodtagere. Til sammenligning er det i serviceeftersynet vurderet, at ca. 1 pct. af lønmodtagerene, der falder bort er en følge af ændringer i beskæftigelsessammensætningen.

5 Lønudviklingen som statistisk mål

I det standardberegnete lønindeks er det lønudviklingen der definerer det statistiske mål, der ønskes belyst gennem det valgte undersøgelsesdesign. For at belyse lønudviklingen renses for ændringer i beskæftigelsessammensætningen, anvendes en

metodik med faste vægte med inspiration fra prisindeks. Målet for hvert kvartal t inden for vægtåret er defineret som:

$$\sum_h w_h^{s_{t_{ref}}} \frac{p_h^t}{p_h^{t_{ref}}},$$

hvor

- $w_h^{s_{t_{ref}}}$ angiver *budgetandelene* for hver homogen gruppe h
- der summeres over *homogene grupper* h , der tilhører en fast, men i notationen uspecificeret gruppe, der enten kan være en branche i en given sektor eller en bestemt hovedarbejdsfunktion.

For hver *homogen gruppe* af lønmodtagere beregnes basisindeks (udviklingen i timeløn siden referencekvartalet), og basisindeksene vægtes derefter sammen til del- og totalindeks. Vægtene udgøres af de årlige lønsummer fra den senest tilgængelige lønstrukturstatistik.

Med denne vægtning måles den rene lønstigning, hvor der er korrigeret for forskydninger i beskæftigelsessammensætningen.

Vægtene skifter på det tidligst mulige tidspunkt, hvilket vil sige så snart de er tilgængelige fra lønstrukturstatistikken - i praksis ved beregningen af 3. kv. Således fås det mest aktuelle vægtgrundlag. Som en del af serviceeftersynet, viste en analyse, at det ikke har nogen særlig betydning for den målte lønudvikling, hvis vægtskiftet flyttes til et kvartal senere.

6 Homogene grupper

De homogene grupper er – som det fremgår ovenfor – centrale. De homogene grupper udgør således grundlaget for beregningen af basisindeks, idet de samler lønmodtagere med samme karakteristika i grupper som grundlag for de enkelte basisindeks.

I det standardberegnete lønindeks er de homogene grupper givet ved disse fire grupperingskategorier:

- Sektor
- Branche
- Arbejdsfunktion
- Aflønningsform (timelønnet eller fastlønnet)

Når lønmodtagerne inddeles i grupperinger baseret på alle kombinationer af disse variable, vil der både være meget store grupper og også en andel af meget små grupper.

Af tabel 1 ses det, at de ti største homogene grupper i hver af de offentlige delsektorer indeholder over halvdelen af det samlede antal lønforhold i delsektoren.

Tabel 1. Koncentration af lønforhold i de største homogene grupper, fordelt på sektor

Sektor	10 største gruppers andel af totale lønforhold i sektoren	
		Pct.
Privat		20,2
Stat		69,8
Kommuner		70,1
Regioner		83,0

Tabel 2 anskueliggør omvendt, hvor mange af de i alt ca. 4.400 homogene grupper, der er i anvendelse ved beregningen for den private sektor.

Tabel 2. Fordeling af homogene grupper og lønforhold i den private sektor på brancher

Branche (10-Grupperingen)	Antal brancher fra 36-grupperingen	Antal homogene grupper	Antal lønforhold (vægtet)	Andel homogene grupper		Andel af lønforhold	
					Pct.		Pct.
2 Industri, råstofindvinding og forsyningsvirksomhed	16	1.162	220.171	41		24	
3 Bygge og anlæg	1	88	57.765	3		6	
4 Handel og transport mv.	3	279	272.312	10		30	
5 Information og kommunikation	3	227	59.297	8		7	
6 Finansiering og forsikring	1	91	64.141	3		7	
7 Ejendomshandel og udlejning	1	96	13.713	3		2	
8 Erhvervsservice	4	367	143.499	13		16	
9 Offentlig administration, undervisning og sundhed	4	320	50.584	11		6	
10 Kultur, fritid og anden service	2	197	28.198	7		3	

Med mange homogene grupper og basisindeks inddeles lønmodtagerne meget findelt, men det vil med stor sandsynlighed også give en stor andel grupper med meget få lønforhold. Små grupper er sensitive over for udskiftninger i de lønforhold, der indgår i beregningerne af gennemsnitslønnen. Samtidig er der større risiko for tomme indeks, hvis indberetninger mangler.

Ved enhver valg af inddeling skal man samtidig have for øje, at grupperingen skal resultere i jobs af samme *kvalitet* i hver gruppe. Det er således et spørgsmål om balance og ydermere vigtigt løbende at overveje, om de homogene grupper på passende vis indfanger forskelle i kvaliteten af det arbejde, der aktuelt udføres på det danske arbejdsmarked.

Ved metodefastlæggelsen er følgende grupperingsvariable blevet analyseret:

- *Sektor og branche* er to vigtige dimensioner for grupperingen af lønmodtagere, da arbejdsopgaver og løndannelse vil være forskellige på tværs af sektorer og i overvejende grad også på tværs af brancher
- *Overenskomster* er af betydning men reflekteres i nogen grad af sektor, branche og ansættelsesform. For store dele af arbejdsmarkedet er der ikke tilgængelige oplysninger om det enkelte jobs tilhørsforhold til overenskomst.

- *Størrelsen på virksomheden*¹, som lønmodtageren er ansat i, kan, særligt på det private arbejdsmarked, også have en betydning for, hvordan lønnen udvikler sig, selvom der ikke nødvendigvis er en direkte sammenhæng mellem størrelse og lønudvikling.
- *Jobfunktionen* og niveauet af færdigheder, som kræves i jobbet, er også af stor betydning. DISCO-koden er en klassifikation heraf.
- *Uddannelse* vil også kunne have påvirkning i og med, at mere uddannelse giver højere specialisering. Men effekten vil til en vis grad være opfanget via DISCO-koden.
- *Erfaring og anciennitet* vil normalt antages at påvirke lønniveau og lønudvikling. Men det kan være problematisk at måle den korrekte og relevante erfaring og anciennitet, da erfaringen i nogen grad vil være knyttet til samme jobtype og sektor. Man kan således godt forestille sig, at en person, der skifter branche eller jobfunktion eller til en ny arbejdsgiver, i nogle tilfælde vil kunne antages at have en mindre grad af relevant erfaring.
- Løndannelsen for *fastlønnede* hhv. *timelønnede* er ret forskellig.

Efter grundige analyser er sektor, branche, jobfunktion (DISCO-kode) og aflønningsform (fast- hhv. timeløn) valgt som de relevante grupperingsvariable ved dannelsen af de homogene grupper. Det er desuden besluttet, at der fastholdes en ensartet struktur på de homogene grupper på tværs af sektorer, så der hvad de homogene grupper i modsætning til situationen før serviceeftersynet er sammenlignelighed på tværs af sektorerne.

7 Opregning til del og totalindeks

Det sidste step i modellen er at opregne fra basisindeks til del- og totalindeks, der er statistikens offentliggørelsesniveau.

Først foretages en opregning af stikprøven, via de korrigerede (kalibrerede) vægte, jf. afsnit 3.2 oven for.

Herefter beregnes gennemsnitstimeløn for hver homogen gruppe, og derved er basisindeks dannet. Basisindeksene sammenvægtes derefter med de relative lønsummer, som de enkelte basisindeks udgør i lønstrukturstatistikken. Dermed er de offentliggjorte del- og totalindeks beregnet.

8 Om match og statistiske modeller

8.1 Mulige modeller

Manualer om prisindeks ser i høj grad på, hvordan et prisindeks konstrueres, hvis der er adgang til perfekte, fejlfri data – eller som minimum adgang til perfekte, fejlfri data fra en stikprøve, hvor alle respondenter svarer til alle tider.

I relation til lønindeks er der imidlertid ikke adgang til perfekte, fejlfri løndata fra alle arbejdssteder i Danmark. Ydermere er der for den private sektor ca. 10 pct. af arbejdsstederne i stikprøven, svarende til ca. 5-6 pct. af lønforholdene, der ikke

¹ Virksomhedsstørrelse anvendes ikke direkte i de homogene grupper, men der tages højde for størrelse i forhold til trækning af og opregning fra stikprøve til population.

indberettes. Det er derfor metodemæssigt nødvendigt at tage højde for dette i forbindelse med valget af statistisk model for lønindekset.

Man kan forestille sig tre forskellige bud på, hvordan man kan konstruere modellen for lønudviklingen iht. det enkelte basisindeks (estimatorer):

(E1): *En direkte estimator for basisindekset, hvor der både i tæller og nævner i estimatoren for basisindekset indgår alle arbejdssteder, der er indberettet for (valget efter serviceeftersynet).*

(E2): *En direkte estimator for basisindekset, hvor der i tæller og nævner i estimatoren for basisindekset kun indgår arbejdssteder, der har indberettet i både referencekvartalet og det aktuelle kvartal (match direkte til referencekvartalet).*

(E3): *En estimator for basisindekset, der er et produkt af estimatorer for kvartalsudviklingerne siden referencekvartalet, og hvor der i hver af estimatorerne indgår præcis de samme arbejdssteder i tæller og nævner (match fra kvartal til kvartal).*

Estimatorerne (E1) og (E2) er baseret på pristeoretiske overvejelser, mens estimatoren (E3) er fremgangsmåden i det standardberegnete lønindeks, der blev udgivet af Danmarks Statistik før serviceeftersynet.

8.2 Valg af model

På baggrund af serviceeftersynet i 2022-2023 er der foretaget et valg af model til det standardberegnete lønindeks, der bedst muligt estimerer lønudviklingen, givet indeksets formål og de data, der er til rådighed for Danmarks Statistik.

En meget vigtig baggrund for valget er karakteren af stikprøven og det ret store bortfald, idet de indsamlede data skal være repræsentative i forhold til populationen. Her er det således ikke i sig selv tilstrækkeligt at have en stor datamængde.

Danmarks Statistiks standardberegnete lønindeks bliver ofte sammenlignet med DA's lønkonjunkturstatistik. Men der er en væsentlig forskel mellem både formålene med indekset og den virkelighed, der forsøges beskrevet via de to produkter.

Hvor formålet med Danmarks Statistiks standardberegnete lønindeks er at beskrive lønudviklingen for alle lønmodtagere i Danmark, beskriver DA formålet med deres konjunkturstatistik (KS) på denne måde: "DA's KonjunkturStatistik skal beskrive udviklingen i prisen på arbejdskraften på DA-området".

Stillet direkte over for hinanden måler Danmarks Statistiks standardberegnete lønindeks lønudviklingen for lønmodtagerne, mens DA's KonjunkturStatistik opgør prisen på arbejdskraft. Og Danmarks Statistik beskriver alle lønmodtagere i alle virksomheder i Danmark, mens populationen for DA's KonjunkturStatistik omfatter den del af virksomhederne, der er medlemmer af DA.

DA lægger ikke i samme grad - som Danmarks Statistik er nødt til det - en stikprøve til grund for statistikken, fordi DA-medlemsvirksomhederne er forpligtede til at indberette til DA.

Samtidig gælder det, at der løbende kan ske udskiftning i DA's population, alene på grund af ind- og udmeldelse af populationen (medlemsskaren). Det betyder, at det for DA i højere grad, end det er for Danmarks Statistik, er relevant at lægge beregningsmodellen an på et krav om matchende virksomheder fra kvartal til kvartal.

For Danmarks Statistiks standardberegnete lønindeks betyder stikprøvemethoden og det markante bortfald derimod, at et krav om matchende virksomheder fra kvartal til kvartal ville give bias, der akkumulerer over tid. I forbindelse med serviceeftersynet er matchkrav derfor opgivet som en af de væsentlige metodeændringer, jf. afsnit 8.1 oven for.

Danmarks Statistik erkender, at opgivelsen af et krav om match fra kvartal til kvartal alt andet lige *kan* give større udsving omkring den langsigtede trend i lønudviklingen. Men det er samtidig Danmarks Statistiks fagligt sikre vurdering, at sådanne udsving for en stikprøvebaseret statistik med et stort bortfald er en nødvendig pris for at undgå bias i form af en for hastigt voksende lønudvikling.