

10. Usikkerhed og fejlsøgning

Forbrugerprisindekset er baseret på en stikprøve af varer og tjenester og derfor behæftet med usikkerhed. Kapitlet indledes derfor med en gennemgang af de væsentligste kilder til usikkerhed og skævhed (bias). Herefter beskrives hvordan de indsamlede prisoplysninger undersøges for fejl.

10.1 Statistisk usikkerhed

Den statistiske usikkerhed kan opdeles i stikprøveusikkerhed og andre kilder til usikkerhed. Stikprøveusikkerheden knytter sig direkte til stikprøven af varer og tjenester, mens andre kilder til usikkerhed vedrører observations- og svarfejl og registreringsfejl.

10.1.1 Stikprøveusikkerhed

Stikprøven af varer og tjenester

Stikprøven af varer og tjenester er ikke baseret på en klassisk stikprøve, hvor sandsynligheden for udtagning svarer til hver enkelt vares eller tjenestes relative betydning. Der kan derfor heller ikke opgøres et enkelt mål for usikkerheden på det samlede prisindeks.

Usikkerhed i vægtgrundlaget

Usikkerhed i vægtgrundlaget påvirker også usikkerheden i prisindekset. Basisindekxsvægtene er opgjort ved at kombinere oplysninger fra forbrugsundersøgelsen og nationalregnskabet og anses som forholdsvis sikre. De detaljerede prisvægte, som anvendes for forretninger eller varegrupper, er typisk behæftet med en større usikkerhed, især for varer eller tjenester med lille omsætning.

Vægtgrundlaget vurderes dog som forholdsvis sikkert og eventuelle fejl skønnes ikke at have indflydelse af betydning for det samlede forbrugerprisindeks. Det er alene i det omfang, prisudviklingen for en bestemt gruppe varer eller tjenester afviger fra den gennemsnitlige prisudvikling, at fejl i vægtgrundlaget vil påvirke indekset. Jo mere ensartet prisudviklingen er på tværs af basisindeksene, jo mindre betydning har vægtgrundlaget; hvis fx alle basisindeks steg med samme procentsats, ville vægtgrundlaget være helt uden betydning.

10.1.2 Andre kilder til usikkerhed

Observationsfejl

Det kan forekomme, at stikprøven kommer til at indeholde elementer som ikke tilhører populationen. Fx kan vægte for forretningers markedsandele indeholde salg til virksomheder, vægte for visse produkter kan indeholde produkter solgt til virksomheder, eller der kan indberettes priser for varer, der (typisk) sælges til erhvervsdrivende. Omvendt kan det ske, at elementer i populationen ikke optræder i stikprøven. Der kan fx være tale om, at der ikke indsamles priser fra visse typer forret-

ninger, fx markeder, postordresalg eller internethandel, og vægten for produkter solgt fra disse forretninger kan være undervurderet.

Svarfejl Der kan opstå svarfejl, hvis en respondent eller prisindsamler opgiver en forkert pris. Dette vil i reglen skyldes misforståelser, hvor der fx tages fejl af hvilke varer der skal indberettes priser for. Det kan også forekomme, at en respondent eller prisindsamler opgiver en forkert pris eller prisen på en vare som kun sælges i ubetydeligt omfang. Non-respons forekommer, hvor en forretning undlader at indsende prisskema eller hvor prisskemaet kun er delvist udfyldt. I disse tilfælde forsøger Danmarks Statistik så vidt muligt at indhente de manglende oplysninger. Det kan også ske at skemaet modtages for sent til at indgå i indeksberegningen.

Registreringsfejl Endelig kan der forekomme fejl i forbindelse med registreringen af priserne i Danmarks Statistik. Der undersøges imidlertid maskinelt for registreringsfejl, som ikke skønnes at udgøre nogen kilde til væsentlige eller systematiske fejl.

Bortfald Ved bortfald forstås priser, som skulle have været med i indeksberegningen, men som af den ene eller anden grund ikke kommer med. Bortfaldet i den månedlige stikprøve af priser er som hovedregel meget beskedent, under 0,5 pct., og skønnes ikke at påvirke det samlede prisindeks.

10.2 Årsager til skævhed i forbrugerprisindekset

Ved skævhed, eller bias, forstås at det opgjorte indeks afviger systematisk fra det, som indekset har til formål at måle. Skævhed er med andre ord forskellen mellem det, som indekset ideelt set skal måle, og det, som indekset faktisk måler.

I forbindelse med forbrugerprisindeks kan der tales om skævhed som følge af flere årsager:

- Substitution mellem varer (formelskævhed)
- Substitution mellem forretninger
- Substitution mellem basisaggregater
- Kvalitetsændringer
- Nye varer og tjenester

Skævhed som følge af substitution mellem varer (formelskævhed)

Med skævhed som følge af substitution mellem varer menes at forbrugerne indenfor det enkelte basisindeks typisk vil substituere fra varer med relativt store prisstigninger, til varer hvis priser relativt er faldet. Hvis basisindekset beregnes som et Carli eller Dutot indeks vil det overvurdere prisstigningernes betydning for husholdningerne. Denne potentielle skævhed er mindre hvis basisindekset beregnes som et Jevons indeks, da der her er taget højde for en hvis substitution. Jævnlig opda-

tering af stikprøven mindsker den potentielle bias som følge af forbrugernes substitution mellem varer og forretninger.

Skævhed som følge af substitution mellem forretninger

Skævhed som følge af substitution mellem forretninger opstår når forbrugerne skifter fra forretninger med høje priser til forretninger med lavere priser. Når der medtages en ny forretning i stikprøven får det først betydning for indekset, når der er indhentet priser fra forretningen i to måneder. Medtagning af fx discountbutikker vil således ikke i sig selv påvirke indekset, selvom det betyder, at husholdningerne i gennemsnit kan købe visse varer billigere. Det får kun betydning indirekte, hvis det medfører at forretninger, som allerede er på markedet, sænker deres priser. Hvis en del af prisforskellen skyldes en kvalitetsforskel (fx ringere service og længere ventetid ved kassen), kan der desuden argumenteres for, at (hele) prisforskellen heller ikke bør medtages i indekset.

Skævhed som følge af substitution mellem basisaggregater

Skævhed som følge af substitution mellem basisaggregater svarer i princippet til formelskævhed, men på et mere aggregeret niveau. De aggregerede prisindeks beregnes ved at sammenveje basisindeksene med et sæt af faste udgiftsvægte. Hvis husholdningerne substituerer fra basisaggregater med relative prisstigninger til basisaggregater med relative prisfald vil indekset derfor overvurdere prisstigningernes betydning for husholdningerne, der typisk ikke holder en fast varekurv. For et inflationsindeks kan der ikke på helt samme måde tales om substitutionsbias, da varekurven her netop bør holdes konstant.

På den anden side giver det ikke mening at holde vægtgrundlaget konstant over lange perioder. Med tiden vil det blive mere og mere misvisende i forhold til forbrugets faktiske sammensætning og indekset vil blive skævt. Vægtgrundlaget bør derfor opdateres med jævne mellemrum således at det holdes repræsentativt i forhold til udviklingen i forbrugets sammensætning.

Skævhed som følge af kvalitetsændringer

Der opstår skævhed som følge af kvalitetsændringer hvis prisændringer, som helt eller delvist skyldes kvalitetsændringer, indregnes i prisindekset. Der bør derfor så vidt muligt korrigeres for værdien af kvalitetsændringer. Generelt gælder, at hvis værdien af kvalitetsforbedringer undervurderes eller helt ignoreres vil det medføre, at prisindekset over tid overvurderer den reelle prisstigning. Det betyder blandt andet, at hvis tidsserier i løbende priser deflateres med et indeks, der overvurderer prisudviklingen, undervurderes de reale vækstrater.

I Boskin (1996) estimeres en bias opad på den årlige vækstrate i det amerikanske forbrugerprisindeks på 0,6 pct. point som følge af manglende hensyntagen til kvalitetsforbedringer.¹ Forskellige internationale studier fra 1993 og frem peger på en bias opad på 0,2 – 0,5 pct. point i forhold til den årlige vækstrate på grund af manglende korrektion for kvalitetsforbedringer (Statistics New Zealand (1997)).

¹ Estimatet dækker både bias i forbindelse med manglende hensyntagen til kvalitetsændringer og bias i forbindelse med introduktion af nye produkter på markedet.

Skævhed som følge af fremkomst af nye varer

Stikprøven af varer og priser opdateres løbende, men af praktiske grunde ofte med en vis tidsforskydning. Det betyder, at nye produkter ofte ikke indgår i beregningen af indekset i deres første levetid, og ikke før der foreligger priser for to på hinanden følgende måneder. I begyndelsen af et produkts levetid er det desuden ofte ikke muligt at få information om omsætningen. Som for formelskævhed gælder også her, at jævnlig opdatering af stikprøven mindsker den potentielle skævhed.

10.3 Fejlsøgning

Formålet med fejlsøgningsproceduren er at identificere og rette fejl i de indsamlede prisoplysninger. Fejlsøgningen foretages hver måned i følgende trin:

- A. Manuel fejlsøgning af indkomne prisoplysninger
- B. Kontrol for konsistens og logiske fejl
- C. Statistisk fejlsøgning
- D. Fejlsøgning af aggregerede prisindeks

A. Manuel fejlsøgning af indkomne prisoplysninger

Skemaer tjekkes

Når skemaerne modtages foretages en gennemgang med henblik på at finde åbenlyse fejl og mangler, fx manglende udfyldelse af skemaet eller skemaer, der ikke kan læses. Eventuelt kontaktes respondenter for en nærmere forklaring.

B. Kontrol for konsistens og logiske fejl

Konsistens-kontrol

Der foretages en række maskinelle kontroller af oplysninger vedrørende respondenter og indberetningshyppighed mv. Dernæst foretages følgende kontroller:

- Alle varer og tjenester undersøges for prismatch (dvs. at der er en pris både i forrige og aktuelle periode)
- Alle priser undersøges for decimalfejl
- Alle respondenter undersøges for, om der har været mindst én prisændring i de sidste 12 måneder ("inlier"-test)
- Vægte undersøges for sumrestriktioner

Observationer på fejllister for disse kontroller undersøges ved hjælp af indberetningsskema eller ved telefonisk kontakt til respondenter.

C. Statistisk fejlsøgning

HB-metoden

For alle produktgrupper med fem eller flere prisændringer identificeres ekstremer ved hjælp af Hidioglou-Berthelot metoden (HB-metoden). Her vurderes prisændringen for hver enkelt vare i forhold til et acceptinterval beregnet på grundlag af medianprisændringen for alle varer i samme produktgruppe.

<i>Fast testgrænse ved få observationer</i>	For produktgrupper med mindre end fem prisændringer identificeres ekstremer som ændringer større end 10 pct. HB-metoden anvendes ikke, hvor der er færre end fem observationer.
<i>Ekstremer tjekkes efter indflydelse</i>	De fundne ekstremeres effekt på de månedlige basisindeks beregnes, og der dannes en fejltabel hvor ekstremerne listes efter størrelsen af denne effekt. Alle ekstremer, der medfører en ændring på 1% eller derover i det månedlige basisindeks undersøges ved tjek af indberetningsskema eller ved kontakt til respondenterne.

D. Fejlsøgning af aggregerede prisindeks

<i>Forbruger- og nettoprisindeks sammenlignes</i>	Prisindeksene beregnes på grundlag af det fejlsøgte og rettede datasæt. Efter beregningen tjekkes delindeks manuelt for udsædvanlige ændringer og der foretages et maskinelt tjek af, om basisindeks i forbruger- og nettoprisindekset har udviklet sig i samme retning.
<i>Sammenligning med andre kilder</i>	Der foretages desuden validering ud fra eksterne kilder. Fx tjekkes observerede priser for benzin, diesel og fyringsolie med tal fra andre kilder. Derudover vurderes prisændringer ud fra kendskab til sæsonbestemte prisudsving mv.

10.3.1 Statistisk fejlsøgning

Den statistiske fejlsøgningsmetode er udviklet af Hidioglou & Berthelot (1986). Metoden anvendes kun for de varer og tjenester, hvor der er en prisændring fra forrige til aktuel måned. Varer, hvor prisen er uændret fra forrige til aktuel måned, indgår ikke i undersøgelsen. Desuden anvendes metoden kun for produktgrupper hvor der er mindst fem prisændringer, da den er mindre stærk med færre observationer.

1. trin Første trin består i at der dannes prisrelativer for alle varer og tjenester, hvor der er en prisændring fra forrige til aktuel måned:

$$r^i = \frac{P_t^i}{P_{t-1}^i}$$

2. trin For hver produktgruppe findes medianværdien r_M af de prisrelativer der indgår i gruppen. Medianen er den midterste værdi i en ordnet talrække. Hvis talrækken består af et lige antal observationer, er medianen gennemsnittet af de to midterste observationer. Medianværdien påvirkes i modsætning til middelværdien ikke af ekstreme værdier.
3. trin For at sikre at eventuelle ekstremer identificeres med samme sandsynlighed i begge ender af fordelingen, foretages for hver produktgruppe en transformation af de individuelle prisrelativer:

$$s^i = \begin{cases} 1 - \frac{r_M}{r^i} & \text{for } 0 < r^i < r_M \\ \frac{r^i}{r_M} - 1 & \text{for } r^i \geq r_M \end{cases}$$

4. *trin* Variablen s^i er niveaauafhængig. For at muliggøre at prisniveauet kan spille en rolle i fejlsøgningen foretages en transformation til en ny variabel, E :

$$E^i = s^i \cdot \left(\max \{ p_{t-1}^i, p_t^i \} \right)^U, \quad 0 \leq U \leq 1$$

E^i beregnes som s^i ganget med den største af forrige og aktuel måneds priser opløftet i potensen U . U bestemmer i hvilken grad, prisniveauet har indflydelse på det acceptinterval, der skal dannes, fx om en ændring fra 10 til 11 kr. skal behandles på samme måde som en ændring fra 100 til 110 kr. Jo større U , jo større indflydelse har prisniveauet. Hvis $U = 0$ har prisniveauet ingen betydning.

5. *trin* For hver produktgruppe findes medianen, E_M , og første og tredje kvartil, henholdsvis E_{Q1} og E_{Q3} , af E^i 'erne. Idet A er en konstant defineres så følgende værdier

$$d_{Q1} = \max \{ E_M - E_{Q1}, |AE_M| \}$$

$$d_{Q3} = \max \{ E_{Q3} - E_M, |AE_M| \}$$

Størrelsen $|AE_M|$ anvendes til at sikre en minimumsbredde på acceptintervallet. Vælges en lav værdi for A øges sandsynligheden for, at det er $(E_M - E_{Q1})$ eller $(E_{Q3} - E_M)$, der kommer til at bestemme d_{Q1} og d_{Q3} . Sættes A fx til 0,05, vil $|AE_M|$ være ganske lille, og der kræves derfor kun en relativ lille spredning i fordelingen af E^i 'erne for at det er $(E_M - E_{Q1})$ eller $(E_{Q3} - E_M)$, der bestemmer d_{Q1} og d_{Q3} . Hvis spredningen i E^i 'erne bliver meget lille, vil det være $|A \cdot E_M|$, der bestemmer d_{Q1} og d_{Q3} .

6. *trin* Det endelige acceptinterval defineres herefter som

$$\{ E_M - Cd_{Q1}; E_M + Cd_{Q3} \}$$

Hvor $(E_M - Cd_{Q1})$ er den nedre grænse i intervallet og $(E_M + Cd_{Q3})$ den øvre grænse. Konstanten C kan i princippet sættes vilkårligt; jo større C , jo bredere bliver acceptintervallet, og jo færre ekstremer og potentielle fejl vil blive identificeret. Der benyttes i 2004 følgende parameter værdier: $U = 0,4$, $A = 0,05$ og $C = 5$.

Selvom HB-metoden identificerer en prisændring som ekstrem, er der ikke nødvendigvis tale om en fejl. Ekstremer defineres som prisændringer, der ligger udenfor det beregnede acceptinterval. Ekstremer kan derfor være store prisstigninger eller prisfald, men også små prisændringer, hvis der generelt er tale om stigende eller faldende priser.