2006/74

Notater

Li-Chun Zhang

Prisindeksberegninger

Notater

Statistisk sentralbyrå

Statistics Norway

Stabsavdeling/Seksjon for statistiske standarder og metoder

Innhold

1 Innledning **3**
1.1 Hva er en prisindeks? 3
1.2 Problem med målenhet for tolkning av prisindekser 3
1.3 Problem med ideelle indekser 3
1.4 Problem med utvalgsteori for prisindekser 4
1.5 Disposisjon 4

2 Direkte indeks **5**
2.1 Fastvektsindeks 5
2.1.1 Referanseperioder 5
2.1.2 Aggregering 5
2.1.3 Prisrelativ og elementær indeks 5
2.1.4 Vekt 6
2.1.5 L-indeks 6
2.1.6 P-indeks 7
2.1.7 Sammenheng mellom L-indeks og P-indeks 7
2.1.8 Eksempel: Prisindeks godstransport på vei 8
2.2 Statistisk teori 9
2.2.1 Modeller for elementære indekser 9
2.2.2 Hva måler man med en modellvarians? 10
2.2.3 Direkte indeks 10
2.2.4 Forhold til indifferensindeks 10
2.2.5 Eksempel: Prisindeks godstransport på vei (fortsettelse) 12
2.2.6 Klynger av prisobservasjoner 12
2.3 Variansestimering 12
2.3.1 Usikkerhetsdisaggregering 12
2.3.2 Robust variansestimering 13
2.3.3 Variansimputering 14

3 Kjedeindeks **16**
3.1 Endring i basisperiode for pris 16
3.1.1 Elementære indekser 16
3.1.2 Aggregering av elementære indekser 16
3.1.3 Teoretisk kjedeindeks 17
3.2 Endring i basisperiode for vekt 18
3.3 Endring i basisperioder for vekt og pris 19
3.4 Endring i indeksreferanseperiode: Renormering 19
3.4.1 Renormering via separat skalering 19
3.4.2 Renormering over flere perioder 21
3.5 Eksempler 21
3.5.1 Kjeding for KPI 21
3.5.2 Kjeding for PIF i 2000 22

4 Hedonisk indeks **24**
4.1 Tradisjonelle ikke-hedoniske metoder 24
4.2 Karakteristikk indeks 25
4.2.1 Hedonisk funksjon 25
4.2.2 Vare og karakteristikk 26
4.2.3 Indeksformel 26
4.2.4 Variansestimering 27
4.3 Dummy-variabel indeks 28
4.3.1 Hedonisk funksjon og indeksformel 28
4.3.2 Statistisk teori 28
4.4 Indekser basert på imputering 30
4.4.1 Metoder 30
4.4.2 Bootstrap variansestimering 31
4.5 Eksempel: Boligprisindeks 32
De sist utgitte publikasjonene i serien Notater 36

1 Innledning

Prisindeksberegning har vært en viktig del av økonomisk statistikkproduksjon i godt over et hundred år. Til tross for den lange historie må det såkalte problem om indekstall ("the index-number problem", Frisch, 1936) fremdeles sies å være uløst. Like fullt er det nødvendig for en statistikkprodusent som SSB å ha et klart *statistisk metodisk* grunnlag for praksis, og ikke minst en standard for begreper og ordbruk. Forhåpentligvis kan dette notat bidra til det.

1.1 Hva er en prisindeks?

En `prisindeks` er et tall som sammenfatter den *relative* prisforskjell mellom to mengder av varer (eller tjenester) som har ingen felles fysisk målenhet. Det er ikke prisnivået men forholdet mellom to sett av priser som måles med en prisindeks. Vanligvis dreier det seg om prisendringer over tid. Men sammenligningen kan like godt gå på tvers av rom eller langs andre dimensjoner. Som regel består ikke de to omtalte mengder av samme individe varer. Gamle varer byttes ut og nye varer settes inn. Det forekommer også stadig endringer i en vare, som ikke nødvendigvis er direkte merkebare. I tillegg er varene fysisk forskjellige fra hverandre siden de mangler en felles målenhet. Så hva betyr et prisindekstall?

1.2 Problem med målenhet for tolkning av prisindekser

"The index-number problem arises whenever we want a quantitative expression for a *complex* that is made up of individual measurements for which no common *physical* unit exists." (Frisch, 1936). Dette gjør at et hvert begrep om et generelt prisnivå til en hypotetisk konstruksjon, så lenge det er definert for en mengde av varer som ikke har en felles målenhet. Dette er utgangspunktet, og mye av grunnen til at problemet fremdeles er uløst.

Frisch skilte mellom to fremgangsmåter i jakten på den ideelle prisindeks, nemlig, den *atomistiske* og den *funksjonelle*, hvis oversettelser jeg har hentet fra Longva (1971). Under den første nevntes bl.a. stokastiske varianten av Edgeworth (1925) og test-varianten av Fisher (1922). Resten av rapporten handlet om den funksjonelle tilnærming (f.eks. Krönus, 1924), der man forsøkte å utlede en entydig indeksformel ut fra et sett av antatte funksjonelle sammenhenger mellom priser og kvanta. Til tross for dens overveldende inflytelse over tolkningen av prisindeks, har den funksjonelle tilnærming fremdeles liten praktisk betydning.

1.3 Problem med ideelle indekser

I alle de ovennevnte teorier er prisindeksen presentert i form av en ideell indeks. Dvs. hvordan en indeksformel ser ut når alle nødvendige data er tilgjengelige. Det finnes ingen skille mellom en teoretisk indeks som en ukjent parameter, slik en hypotetisk konstruksjon nødvendigvis må være, og en beregnet indeks som et estimat for den teoretiske indeks.

Skillet kom mye senere. F.eks. Allen (1963) påpekte at en beregnet Jevons indeks er gitt som forholdet mellom to geometriske gjennomsnitt av observerte priser, som hver for seg kan betraktes som et estimat for det tilsvarende hypotetiske generelle prisnivå. Det er estimatoren som er stokastisk, ikke parameteren. Med et forsvant vanskeligheten til Edgeworth når han skrev følgende om sitt begrep om prisnivå: "To me the conception appears somewhat indefinite as applied to two or a few articles and without relation to the theory of averages."

En nødevendig del av indeksutvikling er derfor å søke bak enhver *indefinite* ideelle indeksformel for å komme fram til den tilsiktede teoretiske *indeksparameter*. Problemet man står igjen med i neste omgang er hvorvidt det finnes en økonomisk fruktbar tolkning av parameteren.

Den stokastiske tilnærming er vel egnet til det (Theil, 1867; Selvanathan and Prasada Rao, 1994; Clements, Izan, and Selvanathan, 2006). Dessverre er det fremdeles en viss avstand mellom teori og praksis. Delvis skyldes det det tradisjonelle fokus på diverse ideelle indeksformler. Vår fremgansmåte har det samme stokastiske perspektiv, men er annerledes på et viktig punkt. Der andre begynner på toppen, starter vi fra bunnen, dvs. indeksen som beregnes i praksis, og bygger opp den statistiske teori derfra.

1.4 Problem med utvalgsteori for prisindekser

I den siste manual for Konsumprisindeks (KPI) (International Labour Office, 2004, s. 69) anvendes utvalgsteori under følgende ramme: (i) "a *universe* consisting of a finite population of units (e.g. products)", (ii) "one or more *variables* that are defined for each unit in the universe (e.g. price and quantity)", og (iii) "a formula which combines the values of one or more of these variables for all units in the universe into a single value called a *parameter* (e.g. the Laspeyres index)".

Men det er feil å betrakte en bestemt statistikk av den endelige populasjon som parameteren, dvs. den teoretiske prisindeks. Uansett hvilken tolkning man legger i begrepet prisindeks, må den være en teoretisk størrelse og ikke en empirisk størrelse. Anvendelse av utvalgsteori for prisindekser strander dermed i første omgang på skillet mellom den teoretiske indeks og den beregnede indeks, som på et dypere plan skyldes problemet med målenhet. Hadde man hatt en fulltelling av alle varene som omsettes i to perioder, inkl. deres priser og kvanta, ville en bestemt statistikk av disse observasjoner likevel kun ha vært et estimat for parameteren. Man kan aldri sette et likhetstegn mellom de to. Tenk om det fantes én ekstra omsetning av en bestemt vare. Den beregnede indeks ville da ha blitt en smule annerledes siden man har en fulltelling. Skal vi tolke det som en endring i estimatet for den samme teoretiske prisindeks, eller skal vi si at den teoretiske indeks per definisjon er annerledes med eller uten denne ene omsetning?

Begrepsforvirringen som kommer fram i punkt (iii) betyr minst to ting i praksis. For det føste er utvalgsvariansen uegnet som usikkerhetsmål for estimering. Selv i tilfellet fulltelling, der utvalgsvariansen er lik null, finnes det usikkerhet i den beregnede indeks, så lenge den må betraktes som et estimat. For det andre må man ta bruken av sannsynlighetsutvalg av priser med en klype salt. Vanligvis betyr det i det minste et to-trinns utvalg, der lokaler (som f.eks. utsalgssteder) trekkes i det første trinn og varer trekkes i det andre. Vanskelighet knyttes spesielt til det siste trinn. Det er usant at sannsynlighetstrekking nødvendigvis er bedre enn trekking av representante varer. Det er meningsløst å kreve et sannsynlighetsutvalg bare for å kunne beregne utvalgsvarians, siden den ikke er det riktige usikkerhetsmål. Her kan vi legge til at idéen om en veldefinert sannsynlighetsbasert utvalgsplan er en illusjon i de alle fleste tilfeller. En gjennomgang av utvalgsteoretisk avhandling som f.eks. Balk (2005) viser nettopp det.

1.5 Disposisjon

I det som følger skal vi begynne med å se på de såkalte *fastvektsindekser* som beregnes i praksis. Derfra forsøker vi i første omgang å klargjøre deres generelle struktur, noe som gjør det mulig for oss å komme til en statistisk fortolkning av den underliggende teoretiske *direkte* indeks, og dermed grunnlaget for statistisk inferens. I neste omgang videreutvikles den statistiske teori til *kjedeindeks* og *hedonisk indeks*. Diskusjoner om den økonomiske indeksteori kan ikke helt unngås, og er dessuten viktig i enkelte deler av teksten, men de kommer til å bli korte. Bevisst har vi valgt en realtivt fin inndeling mellom avsnittene for at man lettere kan finne fram i teksten.

2 Direkte indeks

2.1 Fastvektsindeks

2.1.1 Referanseperioder

Vi må skille mellom 4 forkjellige `referanseperioder`:

- `vektbasisperiode` (b) der vektene hentes fra.
- `prisbasisperiode` (s) av priser i nevneren av prisrelativer.
- `statistikkperiode` (t) av priser i telleren av prisrelativer.
- `indeksreferanseperiode` (r) der indeksen settes til 1. Prosenttall er vanlig ved presentasjon av indekser. Konvertering skal skje der og da, ikke underveis i beregning.

I beregning av fastvektsindekser antar vi at (b, s, r) holdes fast mens t beveger seg i tid. I tillegg har vi $r = s$ slik at indeksen er lik 1 for $t = s$ per definisjon. Kjeding (eller renormering) er nødvendig dersom $r \neq s$. I notasjonen for dette avsnitt kommer derfor r ikke til å synes.

2.1.2 Aggregering

Aggregering defineres vha. en eller flere *klassifiseringsvariabler*. På det laveste nivå finner vi `elementære` aggregeringer (eller grupper), betegnet med $i = 1, ..., M$. En `høyere` aggregering består av en union av elementære aggregeringer. Ingen høyere aggregering kan gå på tvers av elementære aggregeringer. Ofte lager man en hierarkisk struktur blant aggregeringer, slik at forskjellige høyere aggregeringer på samme nivået ikke overlapper med hverandre, og hver for seg er en union av en eller flere aggregeringer på et nivå under. Betegn med G en vilkårlig aggregering. En elementær aggregering skrives som $G = \{i\}$, mens U brukes for den *totale* aggregering.

2.1.3 Prisrelativ og elementær indeks

Betegn med (ij) den $j$te vare (eller tjeneste) i den $i$te elementære gruppe, for $j = 1, ..., n_i$. Betegn med p_{ij}^t prisen på (ij) ved statistikkperiode t, og p_{ij}^s prisen ved prisbasisperiode. La

$$I_{ij}^{s,t} = p_{ij}^t / p_{ij}^s$$

være et `prisrelativ`. Betegn med $P_i^{s,t}$ den $i$te `elementære` indeks, også kjent som mikroindeks. En elementær indeks beregnes *uten* bruk av vekter. Dette er det første steg i beregningen av fastvektsindeks. De tre mest vanlige formler for elementær indeks er

- `Carli` indeks, eller gjennomsnitt av prisrelativer, dvs.

$$P_i^{s,t} = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{p_{ij}^t}{p_{ij}^s} = \frac{1}{n_i} \sum_{j=1}^{n_i} I_{ij}^{s,t} \tag{1}$$

- `Dutot` indeks, eller forhold mellom to prisgjennomsnitter, dvs.

$$P_i^{s,t} = \frac{\frac{1}{n_i} \sum_{j=1}^{n_i} p_{ij}^t}{\frac{1}{n_i} \sum_{j=1}^{n_i} p_{ij}^s} = \frac{\sum_{j=1}^{n_i} p_{ij}^t}{\sum_{j=1}^{n_i} p_{ij}^s} \tag{2}$$

- `Jevons` indeks, eller geometrisk gjennomsnitt av prisrelativer, dvs.

$$P_i^{s,t} = (\prod_{j=1}^{n_i} \frac{p_{ij}^t}{p_{ij}^s})^{\frac{1}{n_i}} = \exp\{\frac{1}{n_i}\sum_{j=1}^{n_i} \log p_{ij}^t - \frac{1}{n_i}\sum_{j=1}^{n_i} \log p_{ij}^s\} \tag{3}$$

Det er viktig å presisere at det vi kaller for en elementær indeks egentlig er et *estimat* på en ukjent parameter, som kan kalles for den teoretiske elementære indeks. Vi har $P_i^{s,t} = I_{i1}^{s,t}$ hvis $n_i = 1$.

2.1.4 Vekt

Elementære indeksene aggregeres opp til en *totalindeks* vha. vekter, betegnet med w_i^b for elementær gruppe i og vektbasisperiode b. Formelt må vektene være slik at

$$w_i^b > 0 \qquad \text{og} \qquad \sum_{i=1}^{M} w_i^b = 1$$

Vi skal anta at vekten er konstruert med sikt på den teoretiske *verdiandel* på gruppenivå, dvs. $v_i^b / \sum_k v_k^b$, der v_i^b er verdien til den $i$te elementære gruppe i vektbasisperiode b.

I estimering av vektene benytter man ofte observasjoner over flere perioder. F.eks. kan vekten for en månedelig indeks være basert på et eller flere tidligere år, som i tilfellet KPI. Ved å samle flere observasjoner reduserer man estimeringsusikkerheten. Ved å se på verdiandelen over et helt år tar man bort uønskede sesongseffekter. Men man skal ikke ta med for mange perioder, for å unngå utdaterte vekter. Mer om dette når vi kommer til tolkningen av en fastvekts indeks.

I tillegg til en totalindeks beregnes det som regel flere *delindekser* på høyere nivåer. For å aggregere elementære indekser til en delindeks som svarer til aggregering G trenger vi vekt $w_{i(G)}^b$ for $i \in G$. Mellom $w_i^b = w_{i(U)}^b$ og en vilkårlig $w_{i(G)}^b$ gjelder følgende transformasjon

$$w_{i(G)}^b = w_{i(U)}^b / \sum_{k \in G} w_{k(U)}^b$$

Som regel skal vi bruke den forenklede notasjon w_i^b der det ikke skaper forvirringer.

2.1.5 L-indeks

En `L-indeks` med vektbasisperiode b og prisbasisperiode s er gitt ved

$$P^{s,t}(b) = \sum_i w_i^b P_i^{s,t} \tag{4}$$

der $P_i^{s,t}$ er den $i$te elementære indeks og w_i^b er dens vekt. For å tydeliggjøre aggregeringen, betegnet med G, kommer vi noen ganger til å skrive

$$P_G^{s,t}(b) = \sum_{i \in G} w_{i(G)}^b P_i^{s,t} \qquad \text{der} \qquad \sum_{i \in G} w_{i(G)}^b = 1$$

En L-indeks $P^{s,t}(b)$ blir en Laspeyres indeks i tilfellet $b = s$ og $P_i^{s,t} = I_i^{s,t} = p_i^t / p_i^s$, da

$$P^{s,t}(s) = \sum_i \frac{q_i^s p_i^s}{\sum_k q_k^s p_k^s} (\frac{p_i^t}{p_i^s}) = \frac{\sum_i q_i^s p_i^t}{\sum_i q_i^s p_i^s}$$

Ellers er $v_i^b = p_i^b q_i^b$, dvs. verdi som et produkt av pris og kvantum, kun en forestilling. Under denne forestilling har man, i tilfellet $b \neq s$,

$$P^{s,t}(b) = \sum_i \frac{q_i^b p_i^b P_i^{s,t}}{\sum_k q_k^b p_k^b} = \frac{\sum_i q_i^b \tilde{p}_i^{b+t-s}}{\sum_i q_i^b p_i^b} \qquad \text{der} \qquad \tilde{p}_i^{b+t-s} = p_i^b P_i^{s,t}$$

dvs. prisen i periode b justert framover (for $s < t$) med prisendringen mellom periode s og t. Hadde det vært mulig å bruke den faktiske p_i^{b+t-s} ville man ha fått en Laspeyres indeks, der man fastholder kvanta q_i^b i perioden b når man sammenligner prisene mellom b og $b + (t - s)$.

2.1.6 P-indeks

En `P-indeks` med vektbasisperiode b og prisbasisperiode s er gitt ved

$$\wp^{s,t}(b) = \{\sum_i w_i^b (P_i^{s,t})^{-1}\}^{-1} \tag{5}$$

der $P_i^{s,t}$ er den $i$te elementære indeks og w_i^b er dens vekt. En P-indeks blir en Paasche indeks i tilfellet $b = t$ og $P_i^{s,t} = I_i^{s,t} = p_i^t / p_j^s$, da

$$\wp^{s,t}(t) = \{\sum_i \frac{q_i^t p_i^t}{\sum_k q_k^t p_k^t} (\frac{p_i^t}{p_i^s})^{-1}\}^{-1} = \{\frac{\sum_i q_i^t p_i^s}{\sum_i q_i^t p_i^t}\}^{-1} = \frac{\sum_i q_i^t p_i^t}{\sum_i q_i^t p_i^s}$$

Men legg merke til at da er den ikke lenger en fastvektsindeks. Ellers kan man forestille seg at v_i^b faktoriseres som et produkt av pris og kvantum, slik at

$$\wp^{s,t}(b) = \{\sum_i \frac{q_i^b p_i^b / P_i^{s,t}}{\sum_k q_k^b p_k^b}\}^{-1} = \frac{\sum_i q_i^b p_i^b}{\sum_i q_i^b \tilde{p}_i^{b-t+s}} \qquad \text{der} \qquad \tilde{p}_i^{b-t+s} = p_i^b / P_i^{s,t}$$

dvs. prisen i periode b justert bakover (for $s < t$) med prisendringen mellom periode s og t. Hadde det vært mulig å bruke den faktiske p_i^{b-t+s} ville man ha fått en Paasche indeks, der man fastholder kvanta q_i^b i perioden b når man sammenligner prisene mellom $b - (t - s)$ og b.

2.1.7 Sammenheng mellom L-indeks og P-indeks

Både L-indeksen og P-indeksen er fastvektsindekser så lenge b holdes fast mens t beveger seg i tid. Sammenhengen mellom (4) og (5) lar seg fastslå gitt *tidsombyttbar* elementær indeks, dvs.

$$P_i^{s,t} = 1/P_i^{t,s}$$

Både Dutot og Jevons indeksen er tidsombyttbar, men ikke Carli indeksen. Derfor, gitt Dutot eller Jevons indeks, har vi

$$\wp^{s,t}(b) = \{\sum_i w_i^b (P_i^{s,t})^{-1}\}^{-1} = \{\sum_i w_i^b P_i^{t,s}\}^{-1} = 1/P^{t,s}(b)$$

Dvs. P-indeksen mellom (s, t), der prisene i periode s er satt til 1, er lik invers til L-indeksen mellom (t, s), der prisene i periode t er satt til 1. Man finner det samme forhold mellom Laspeyres og Paasche indeksen. Det blir imidlertid galt å kalle (4) for en Laspeyres indeks, eller (5) for en Paasche indeks, noe de åpenbart ikke er.

2.1.8 Eksempel: Prisindeks godstransport på vei

Prisindeks for godstransport på vei Prisindeks for godstransport på vei (NACE 60.24) er basert på en kvartalsvis undersøkelse om priser på transport turer/sendinger. En tur er definert mht. frakttype, kunde, lasttype, mengde og strekning. Utvalget er et årlig roterende panel, som består av 150 - 200 bedrifter med næringskode 60.24, trukket med en sannsynlighet som er proporsjonal med bedriftens omsetning. Hver av disse får tilsendt et skjema, med mulighet til å oppgi pris på 4-6 turer, eller alle turer hvis bedriften har færre enn 4 turer. Vi viser til Dokumentasjonsnotat "Godstransport på vei - kvartalsvis prisindeks" (av J. Melbye og P.-C. Lysaker Torgersrud) for detaljene. Her konsentrerer vi oss om beskrivelsen av beregningsopplegget.

Beskrivelse av beregningsopplegget — Et sammendrag For hver tur blir det beregnet en prisendring i forholdet til det 4. kvartal året før. Ut fra disse beregnes det en bedriftsindeks per frakttype, som er et geometrisk gjennomsnitt av prisendringene med den gitte frakttypen. Deretter beregnes det en delindeks per frakttype for hhv. små, mellomstore og store bedrifter, som et aritmetisk gjennomsnitt av bedriftsindeksene per frakttype innen hver stratum av bedrifter. Det er 6 forskjellige frakttyper, slik at det til sammen er 18 ($= 6 \times 3$) delindekser per frakttype og bedriftsstratum. Delindeksen per frakttype er nå gitt som en veid sum over de 3 bedriftsstrata. Vektene er gitt ved de tilsvarende omsetningsandeler i 2003 basert på Strukturundersøkelse for transport og kommunikasjon 2004. Til slutt beregnes totalindeksen som en veid sum over de 6 delindekser per frakttype, der vektene er gitt ved tilsvarende omsetningsandelene.

Omskriving av beregningsopplegget — En standard Betegn med t det inneværende kvartal. Prisbasisperiode s er det 4. kvartal året før. Vektbasisperiode b er for øyeblikket år 2003. Indeksreferanseperioden $r = s$ for fastvektsberegning. Den elementære gruppe av turer er definert vha. klassifiseringsvariabler frakttype (6 typer), bedriftsstratum (3 strata) og bedrift. Betegn med (ij) den $j$te tur i den $i$te elementære gruppe. For hver tur blir det beregnet et prisrelativ, betegnet med $I_{ij}^{s,t} = p_{ij}^t / p_{ij}^s$. Den $i$te elementære indeks er en Jevons indeks, dvs. et geometrisk gjennomsnitt, basert på $I_{ij}^{s,t}$ for $j = 1, ..., n_i$, der n_i er antallet turer. Fra elementær aggregering oppover har vi hierarkisk følgende aggregeringer: frakttype og bedriftsstratum, betegnet med $h = 1, ..., H$; frakttype; og total aggregering. La W_h^b betegne omsetningsandelen til den $h$te aggregering av frakttype og bedriftsstratum basert på Strukturundersøkelse om vektbasisperioden. La n_h være antallet bedrifter i det tilsvarende bedriftsstratum som har oppgitt tur (eller turer) av den tilsvarende frakttype, betegnet med $i = 1, ..., n_h$. Vekten til den $i$te elementære gruppe er da gitt ved $w_i^b = W_h^b / n_h$. Delindeks per frakttype og bedriftsstratum, per frakttype, og totalindeksen er alle L-indekser. Spesielt er vekten til delindeks per frakttype og bedriftsstratum gitt ved

$$w_{i(h)}^b = w_i^b / (\sum_{k \in h} w_k^b) = (W_h^b / n_h) / W_h^b = 1/n_h$$

Alternativ beregningsmåte Det at $w_{i(h)}^b = 1/n_h$ betyr at man behandler hver bedrift i utvalget som om de representerer en like stor omsetningsandel i populasjonen per frakttype og bedriftsstratum. Er det da like greit å definere den elementære aggregering på frakttype og bedriftsstratums nivå? I så fall er det naturlig å bruke en Jevons indeks som elementær indeks, dvs.

$$P_h^{s,t} = \exp\{(\sum_{i \in h} \sum_{j \in i} \log I_{ij}^{s,t}) / (\sum_{i \in h} n_i)\}$$

Man kan da vurdere å innføre finere inndeling mht. bedriftsomsetning, for å gjøre de elementære grupper mer homogene. Dette er mulig siden det finnes mange flere turer per frakttype og bedriftsstratum, enn per bedrift, frakttype og bedriftsstratum. For å velge mellom de to alternativer trenger vi å avklare de to tilsvarende teoretiske indekser.

2.2 Statistisk teori

Hvilken teoretisk prisindeks estimerer man med en fastvekts indeks? Hva er parameteren?

2.2.1 Modeller for elementære indekser

De tre elementære indekser svarer til den beste lineære forventningsrette estimator (BLUE) under tre bestemte lineære regresjonsmodeller. Solheim (2004) inneholder lignende beskrivelser.

Carli indeks Modellen for Carli indeks (1) er gitt ved

$$I_{ij}^{s,t} = \frac{p_{ij}^t}{p_{ij}^s} = \theta_i + \epsilon_{ij} \qquad \text{der} \quad E(\epsilon_{ij}) = 0 \quad \text{og} \quad V(\epsilon_{ij}) = \sigma_i^2 \quad \text{og} \quad Cov(\epsilon_{ij}, \epsilon_{ik}) = 0 \tag{6}$$

dvs. en *gruppe homogen* modell for prisrelativer, med konstant gruppegjennomsnitt og -varians. Carli indeksen er BLUE'en for θ_i, som da er den teoretisk elementære indeksparameter. Modellen kan omskrives som

$$p_{ij}^t = \theta_i p_{ij}^s + \epsilon_{ij} \qquad \text{der} \quad E(\epsilon_{ij}) = 0 \quad \text{og} \quad V(\epsilon_{ij}) = \sigma_i^2 (p_{ij}^s)^2 \quad \text{og} \quad Cov(\epsilon_{ij}, \epsilon_{ik}) = 0$$

dvs. en *gruppe rate* modell for p_{ij}^t gitt p_{ij}^s, der residualen har en varians som er proporsjonal med $(p_{ij}^s)^2$. Uansett er estimeringsusikkerheten, dvs. variansen til Carli indeks, gitt ved

$$V(P_i^{s,t}) = n_i^{-2} \sum_j (p_{ij}^s)^{-2} V(p_{ij}^t) = \sigma_i^2 / n_i \tag{7}$$

Dutot indeks Modellen for Dutot indeks (2) er gitt ved

$$p_{ij}^t = \theta_i p_{ij}^s + \epsilon_{ij} \qquad \text{der} \quad E(\epsilon_{ij}) = 0 \quad \text{og} \quad V(\epsilon_{ij}) = \sigma_i^2 p_{ij}^s \quad \text{og} \quad CovV(\epsilon_{ij}, \epsilon_{ik}) = 0 \tag{8}$$

dvs. en *gruppe rate* modell med varians proporsjonal med p_{ij}^s. Den eneste forskjell fra modell (6) ligger i variansantagelsen. Dutot indeksen er BLUE'en for θ_i, som er den teoretisk elementære indeksparameter. Variansen til Dutot indeksen under modell (8) er gitt ved

$$V(P_i^{s,t}) = (\sum_{j=1}^{n_i} p_{ij}^s)^{-2} \{\sum_j V(p_{ij}^t)\} = \sigma_i^2 / (\sum_{j=1}^{n_i} p_{ij}^s) \tag{9}$$

Jevons indeks Modellen for Jevons indeks (3) er gitt ved

$$\log I_{ij}^{s,t} = \mu_i + \epsilon_{ij} \quad \text{der} \quad E(\epsilon_{ij}) = 0 \quad \text{og} \quad V(\epsilon_{ij}) = \sigma_i^2 \quad \text{og} \quad Cov(\epsilon_{ij}, \epsilon_{ik}) = 0 \tag{10}$$

dvs. en gruppe homogen modell for logaritmen av prisrelativer. Modellen kan omskrives som

$$p_{ij}^t = p_{ij}^s \exp(\mu_i) \exp(\epsilon_{ij})$$

dvs. en multiplikativ regresjonsmodell for p^t_{ij} gitt p^s_{ij}. Uansett er Jevons indeks BLUE'en for μ_i, mens den teoretisk elementære indeksparameter er

$$\theta_i = \exp(\mu_i) \qquad \text{og} \qquad E(P_i^{s,s+1}) \approx \theta_i$$

I tillegg er variansen til Jevons indeks under modell (10) gitt ved

$$V(P_i^{s,t}) \approx e^{2\mu_i} n_i^{-2} \sum_j V(\log p^t_{ij}) = e^{2\mu_i}\sigma_i^2/n_i = \theta_i^2\sigma_i^2/n_i \tag{11}$$

2.2.2 Hva måler man med en modellvarians?

Hva måler man med modellvariansen (7), (9) eller (11)? Anta at den $i$te elementære gruppe består av to observasjoner, der $p^s_{i1} = p^s_{i2} = 100$ kroner. Anta først $p^t_{i1} = 100$ kroner og $p^t_{i2} = 120$ kroner. Dutot indeksen er da $P_i^{s,t} = 1.1$. Anta nest $p^t_{i1} = 109$ kroner og $p^t_{i2} = 111$ kroner. Dutot indeksen er igjen $P_i^{s,t} = 1.1$. Men det er klart at variasjonen i p^t_{ij}/p^s_{ij} er forskjellig i de to tilfeller. Det er denne variasjon som formidles som usikkerheten i estimeringen. Spesielt er det to ting som påvirker estimeringsusikkerheten: variasjon i enkelte prisobservasjoner p^t_{ij} (eller $\log p^t_{ij}$) gitt p^s_{ij}, samt antallet observasjoner som ligger bak en elementær indeks. Det kan nevnes at i praksis bruker man av og til som basispris et gjennomsnitt av priser over flere perioder. Den har ingen klare variansreduserende effekter. Dog kan den være nyttig i situasjoner der elementære aggregeringer ligger så lavt at man risikerer tomme observasjoner fra tid til annen.

2.2.3 Direkte indeks

Vi har altså 3 forskjellige statistiske modeller (6), (8) og (10) for den teoretiske underliggende prisendring, betegnet med θ_i (eller $\theta_i^{s,t}$), for elementær aggregering $i = 1, ..., M$. Betrakt w_i^b som estimatet for den teoretiske verdiandel til den $i$te elementære gruppe, betegnet med ω_i (eller ω_i^b). Den (teoretiske) `direkte` indeks, dvs. paremeteren, for en L-indeks er da gitt ved

$$\theta = \theta^{s,t}(b) = \sum_i \omega_i^b \theta_i^{s,t} \tag{12}$$

Vi kan betrakte $\{\omega_i^b\}$ som en punktvis sannsynlighetsfunksjon siden $\sum_i \omega_i^b = 1$. *Parameteren θ er da den forventede prisendring fra s til t man møter når én krone skal brukes, og sannsynligheten for at den skal brukes på den $i$te gruppe av varer er lik ω_i^b.* Parameteren for en delindeks $P_G^{s,t}(b)$ blir da den *betingede* forventning gitt at kronen skal brukes på varer under aggregering G, siden $w^b_{i(G)}$ svarer til denne betingede sannsynlighetsfordeling. På lignende måte er *invers* til parameteren for en P-indeks, gitt tidsombyttbar $P_i^{s,t}$, lik den forventede prisendring fra t til s

$$\theta^{-1} = \theta^{t,s}(b) = \sum_i \omega_i^b \theta_i^{t,s} \tag{13}$$

Forskjellen mellom L- og P-indeksen skyldes i hvilken retning man velger å betrakte endringen.

2.2.4 Forhold til indifferensindeks

Vi har kommet fram til en statistisk fortolkning av fastvektsindeksene, som i sin tur danner grunnlaget for statistisk inferens. Hva med det økonomiske spørsmål om indekstall? Hva er f.eks. sammenhengen mellom en direkte indeks (12) og en indifferensdefinert prisindeks?

En vare per elementær gruppe Anta at hver elementær gruppe betår av kun én vare, som representerer et bestemt *behov*. En bestemt sammensetning av kvanta til forksjellige varer gir en bestemt *nytte*[1]. Samme nytten kan oppnås ved forskjellige sammensetninger av varer. Dette er et spørsmål om *smak*. Under den funksjonelle tilnærming betraktes prisen som en gitt konstant, betegnet med $(p_i^s, ..., p_M^s)$ og $(p_1^t, ..., p_M^t)$. Den tilsvarende entydige sammensetning av kvanta til varer avgjøres *funksjonelt* av smak og rasjonalitet, betegnet med $(q_i^s, ..., q_M^s)$ og $(q_1^t, ..., q_M^t)$. Det å være rasjonell her betyr at man alltid søker å minimere kostnad gitt nytten. Når nytten er den samme ved begge tidspunkter er den tilsvarende *indifferensindeks* gitt ved

$$(\sum_i q_i^t p_i^t)/(\sum_i q_i^s p_i^s)$$

Legg merke til at hvor mye forskjellige varer faktisk omsettes i et bestemt marked over en bestemt periode er hverken interessant eller relevant under den funksjonelle tilnærming.

I realiteten vet vi ingen detaljer om den funksjonelle sammenheng mellom kvanta og priser og nytte. Anta at $\omega_i^b = v_i^b / \sum_i v_i^b$, der $v_i^b = q_i^b p_i^b$. Anta at q_i^b er funksjonelt bestemt av $(p_1^b, ..., p_M^b)$ gitt nytten. Her kan b tolkes som en *smaksindikator* som, sammen med rasjonalitet, har resultert i $(q_1^b, ..., q_M^b)$. Da kan parameteren i (12), dvs. direkte indeksen, skrives som

$$\theta = (\sum_i q_i^b p_i^b \theta_i^{s,t})/(\sum_i q_i^b p_i^b)$$

Dvs. et *teoretisk* kostnadsforhold gitt basisprisene $(p_1^b, ..., p_M^b)$ og den teoretiske prisutvikling $\theta_i^{s,t}$ (mellom s og t) når *handelskurven* $(q_i^b, ..., q_M^b)$ holdes fast. Den er ikke en indifferensindeks. Til det må man anta at $(q_1^b, ..., q_M^b)$ er identiske med kvanta som er funksjonelt bestemt av oppdaterte prisene $(p_1^b \theta_1^{s,t}, ..., p_M^b \theta_M^{s,t})$ gitt nytten. Som sagt vet vi ingenting om det. Men θ gir en øvre grense for indifferensindeksen dersom $(q_1^b, ..., q_M^b)$ gir samme nytten på begge tidspunkter, siden det rasjonelle valg basert på oppdaterte prisene ikke kan koste mer. Legg merke til at $b = s$ er hverken nødvendig eller tilstrekkelig for å gjøre θ til en indifferensindeks. Men å holde b nærmest mulig s gjør trolig både direkte indeksen og indifferensindeksen mer oppdaterte og mer relevante i forhold til opplevelsen av prisutviklinger.

Generelt Forholdet mellom direkte- og indifferensindeksen blir enda mer uoversiktig når det i realitet finnes flere enn én vare innen hver elementær gruppe. Kan man i det hele tatt snakke om en funksjonell sammenheng mellom priser og kvanta på elementære aggregeringsnivået? Hva er da q_i gitt varer uten en felles målenhet? Også tolkningen av $\theta_i^{s,t}$ kan diskuteres. Frisch (1936) skrev "We cannot assume that the 'monetary factor' will manifest itself as a proportional change of all prices." Klart hvis det handler om *alle* priser. Men kan vi anta én underliggende prisendring hvis vi begrenser oss til *noen* varer? Kan vi snakke om en teoretisk pris for samme vare som selges i forskjellige lokaler, selv om den økonomiske teori ikke tillater slik variasjon i et perfekt marked? Kan vi forestille oss to varer som er så like med hverandre at en typisk forbruker er likegyldig overfor den ene eller den andre, så lenge det er samme pris på begge to? Hvor langt kan vi strekke begrepet om slike *indifferensvarer*? Uansett er det klart at *man bør prøve å komme fram til varer som har en homogen prisutvikling i definisjonen av elementære aggregeringer.*

[1] Her antar man at nytte kun avhenger av kvanta, men ikke priser. Dette skyldes at man antar at det til enhver tid finnes kun én pris på en bestemt vare i markedet. Det samme får man også ved å innføre én teoretisk pris på en bestemt vare. I tillegg er resonnementet som følger gyldig selv når nytten avhenger av både kvanta og priser.

2.2.5 Eksempel: Prisindeks godstransport på vei (fortsettelse)

I Eksempelet med GPV ovenfor nevnte vi to alternative beregningsmåter. Den statistiske fortolkning av direkte indeks kan hjelpe oss med å komme fram til en avgjørelse. Det nåværende beregningsopplegg svarer til modell (10) for prisen per bedrift, frakttype og bedriftsstratum. Samtidig forutsetter det at $w^b_{i(h)} = 1/n_h$ er et fornuftig estimat på omsetningsandelen til den $i$te bedrift gitt frakttype og bedriftsstratum, noe den åpenbart ikke er. Det alternative opplegg svarer til modell (10) for prisen per frakttype og bedriftsstraum, uansett hvilken bedrift den kommer fra. Vekten W^b_h er per definisjon et fornuftig estimat på ω^b_h, når $\{h\}$ svarer til en elementær aggregering per frakttype og bedriftsstraum. Det alternative beregningsoppplegg synes å bedre oppfylle de statistiske antagelser. Som sagt kan man vurdere å innføre finere stratumsinndeling på bedrifter for å gjøre antagelsen om homogenitet enda mer plausibel.

2.2.6 Klynger av prisobservasjoner

Det er klart at modell (6), (8) og (10) inneholder en del forenklinger. Dette gjelder spesielt korrelasjoner på forskjellige nivåer som skyldes at prisobservasjonene samles inn i klynger.

- Det er f.eks. vanlig at priser som er samlet inn fra et lokale klassifiseres under forskjellige elementære grupper, siden det er sjelden at det der finnes kun én type varer, og man sparer kostnader ved å samle inn flere typer priser fra et lokale som med i undersøkelsen. Det er i så fall ikke uten videre gitt at elementære indekser er uavhengige av hverandre.
- Av samme grunnen er det også vanlig at man samler inn flere priser fra det samme lokale, som hører til under samme elementære aggregeringen, slik at observasjonene heller ikke er uavhengige innenfor en elementær aggregering. F.eks. Valliant (1992) studerer variansestimering der modell (6) er utvidet til å inkludere denne type korrelasjon.
- Klyngeeffekt kan også oppstå på en høyere aggregering. F.eks. prisrelativer fra bedrifter innenfor en bestemt type virkesomhet kan tenkes å være korrelerte.

Imidlertid finnes det ikke en generell formulering som greier å ta hensyn til alle typer korrelasjoner. Hvor mye disse eventuellt påvirker variansen må vurderes i hvert enkelt tilfelle.

Et annet spørsmål som da oppstår er følgende. Man kan f.eks. gå fra en gruppe ratemodell, som gir opphav til Dutot indeksen, til en hierarkisk blandet modell, for å bedre ivareta korrelasjonen blant prisobservasjonene innen den samme elementære gruppe. Men da bør man også revurdere selve estimatet, dvs. fastvektsindeksen. Her skal vi ikke forfølge dette spor lenger.

2.3 Variansestimering

2.3.1 Usikkerhetsdisaggregering

Først skal variansen til en fastvektsindeks brytes ned til variansene til elementære indekser. Legg merke til at basisvekten her er konstant per definisjon. Vi har derfor, i tilfellet L-indekser,

$$V(P^{s,t}(b)) = \sum_i a_i V(P^{s,t}_i) \qquad \text{og} \qquad a_i = (w^b_i)^2 \tag{14}$$

For å komme til et uttrykk for P-indekser bruker vi en tilnærming basert på Taylor utviklingen

$$Z^{-1} \approx E(Z)^{-1} - E(Z)^{-2}\{Z - E(Z))\} \qquad \Rightarrow \qquad V(Z^{-1}) \approx E(Z)^{-4} V(Z)$$

for en tilfeldig variabel Z. Vi har da, i tilfellet P-indekser,

$$\begin{aligned} V(\wp^{s,t}(b)) &\approx E(\wp^{s,t}(b))^4\{\sum_i (w_i^b)^2 E(P_i^{s,t})^{-4} V(P_i^{s,t})\} \\ &= \sum_i a_i V(P_i^{s,t}) \qquad \text{der} \qquad a_i = (w_i^t)^2 E(\wp^{s,t}(b))^4 / E(P_i^{s,t})^4 \end{aligned} \tag{15}$$

2.3.2 Robust variansestimering

En direkte variansestimator fås ved å erstatte (σ_i^2, θ_i) med $(\hat{\sigma}_i^2, P_i^{s,t})$ der de forkommer i variansformelene (7), (9) og (11). En slik variansestimator er imidlertid lite robust. Selv om en antatt variansstruktur kan brukes til å motivere den tilsvarende elementære indeks, så behøver ikke den å gi en riktig beskrivelse av de faktiske prisdata. Vi kan estimere variansen på en måte som er uavhengig av disse variansantagelser. Metoden vi bruker er et spesielt tilfelle av såkalt sandwich variansestimering (e.g. Valliant, Dorfman, and Royall, 2000, Kapittel 5).

Carli indeks Det første uttrykk i variansformel (7) viser at generelt trenger vi å estimere $V(I_{ij}^{s,t}) = \sigma_{ij}^2$. Det andre uttrykk følger av antagelsen $\sigma_{ij}^2 = \sigma_i^2$. Den trengs for at Carli indeksen kan betraktes som BLUE under modell (6). La

$$e_{ij} = I_{ij}^{s,t} - \hat{\theta}_i = I_{ij}^{s,t} - P_i^{s,t} = I_{ij}^{s,t} - n_i^{-1} \sum_k I_{ik}^{s,t}$$

slik at

$$E(e_{ij}^2) = (1 - \frac{1}{n_i})^2 \sigma_{ij}^2 + \frac{1}{n_i^2} \sum_{k \neq j} \sigma_{ik}^2$$

Vi estimerer variansen til en Carli indeks ved

$$v_i = \widehat{V}(P_i^{s,t}) = n_i^{-2} \frac{n_i}{n_i - 1} \sum_j e_{ij}^2 = \frac{\sum_j e_{ij}^2}{n_i(n_i - 1)} \tag{16}$$

slik at

$$\begin{aligned} E(v_i) &= \frac{1}{n_i^2} \sum_j (1 - \frac{1}{n_i})^{-1} E(e_{ij}^2) = \frac{1}{n_i^2} \{\sum_j (1 - \frac{1}{n_i}) \sigma_{ij}^2 + \frac{1}{n_i(n_i - 1)} \sum_j \sum_{k \neq j} \sigma_{ik}^2\} \\ &= \frac{1}{n_i^2} \{(1 - \frac{1}{n_i})(\sum_j \sigma_{ij}^2) + \frac{n_i - 1}{n_i(n_i - 1)} (\sum_k \sigma_{ik}^2)\} = \frac{1}{n_i^2} \sum_j \sigma_{ij}^2 = V(P_i^{s,t}) \end{aligned}$$

Mao. v_i er forventningsrett under modell (6) selv uten antagelsen $\sigma_{ij}^2 = \sigma_i^2$.

Dutot indeks Det første uttrykk i variansformel (9) viser at vi trenger å estimere $V(p_{ij}^t) = \sigma_{ij}^2$. Det andre uttrykk følger av antagelsen $\sigma_{ij}^2 = \sigma_i^2 p_{ij}^s$. La

$$e_{ij} = p_{ij}^t - \hat{\theta}_i p_{ij}^s = p_{ij}^t - P_i^{s,t} p_{ij}^s = p_{ij}^t - \frac{\sum_k p_{ik}^t}{\sum_k p_{ik}^s} p_{ij}^s = (1 - \frac{p_{ij}^s}{\sum_k p_{ik}^s}) p_{ij}^t + \frac{p_{ij}^s}{\sum_k p_{ik}^s} \sum_{l \neq j} p_{il}^t$$

slik at

$$E(e_{ij}^2) = V(e_{ij}) = (1 - \frac{p_{ij}^s}{\sum_k p_{ik}^s})^2 \sigma_{ij}^2 + (\frac{p_{ij}^s}{\sum_k p_{ik}^s})^2 \sum_{l \neq j} \sigma_{il}^2$$

Vi estimerer variansen til en Dutot indeks ved

$$v_i = \widehat{V}(P_i^{s,t}) = (\sum_k p_{ik}^s)^{-2} \sum_j (1 - \frac{p_{ij}^s}{\sum_k p_{ik}^s})^{-1} e_{ij}^2 \tag{17}$$

som er tilnærmet forventningsrett under modell (8) uavhengig av antagelsen $\sigma_{ij}^2 = \sigma_i^2 p_{ij}^s$.

Jevons indeks Det første uttrykk i variansformel (11) viser at vi trenger å estimere $V(\log I_{ij}^{s,t}) = \sigma_{ij}^2$. Det andre uttrykk følger av antagelsen $\sigma_{ij}^2 = \sigma_i^2$. La

$$e_{ij} = \log I_{ij}^{s,t} - \hat{\mu}_i = \log I_{ij}^{s,t} - n_i^{-1} \sum_k \log I_{ik}^{s,t}$$

En tilnærmet forventningsrett variansestimator for en Jevons indeks er da gitt ved

$$v_i = \widehat{V}(P_i^{s,t}) = (P_i^{s,t})^2 \frac{\sum_j e_{ij}^2}{n_i(n_i - 1)} \tag{18}$$

2.3.3 Variansimputering

I en del situasjoner kan elementære aggregeringer ligge så lavt at mange elementære indekser er basert på kun én observasjon. Variansestimatene (16) - (18) er udefinerte i slike tilfeller. Betrakt følgende tre alternativer for variansimputering.

Alternativ I Sett $v_i = 0$. Metoden resulterer i en underestimering av variansen.

Alternativ II Imputer e_{ij}^* (der $j = 1 = n_i$) og, deretter, basert på (6), (8) og (10),

$$v_i^* = \begin{cases} (e_{i1}^*)^2 & \text{i tilfellet Carli indeks} \\ (e_{i1}^*)^2/(p_{i1}^s)^2 & \text{i tilfellet Dutot indeks} \\ (e_{i1}^*)^2(P_i^{s,t})^2 & \text{i tilfellet Jevons indeks} \end{cases}$$

Det er selvsagt flere måter som kan tenkes når det gjelder imputering av e_i^*. F. eks.

$$e_{i1}^* = \begin{cases} I_{i1}^{s,t} - n^{-1} \sum_{l,j} I_{lj}^{s,t} & \text{i tilfellet Carli indeks} \\ p_{i1}^t - p_{i1}^s(\sum_{l,j} p_{lj}^t)/(\sum_{l,j} p_{lj}^s) & \text{i tilfellet Dutot indeks} \\ \log(I_{i1}^{s,t}) - n^{-1} \sum_{l,j} \log(I_{lj}^{s,t}) & \text{i tilfellet Jevons indeks} \end{cases}$$

Mao. den imputerte e_{i1}^* er lik e_{i1} i tilfellet der det finnes kun én elementær gruppe og $P_i^{s,t} = P^{s,t}$. Derfor inneholder e_{i1}^* også variasjonen mellom grupper, som f. eks. i tilfellet Carli indeks,

$$e_{ij}^* = (I_{i1}^{s,t} - n_i^{-1} \sum_j I_{ij}^{s,t}) + (n_i^{-1} \sum_j I_{ij}^{s,t} - n^{-1} \sum_{l,j} I_{lj}^{s,t})$$

Følgelig har den imputerte v_i^* en positiv skjevhet, slik at variansen blir overestimert.

Alternativ III Imputer v_i^* direkte. Sammenligning mellom (7) og (16) viser at $\hat{\sigma}_i^2 = n_i v_i$ kan betraktes som et estimat for σ_i^2 når $n_i > 1$. En mulighet i tilfellet Carli indeks er derfor å sette

$$v_i^* = (\sum_l I_{n_l>1} n_l v_l)/(\sum_l I_{n_l>1}) \qquad \text{for} \quad n_i = 1 \tag{19}$$

der $I_{n_l>1} = 1$ hvis $n_l > 1$ og 0 ellers. Dvs. gjennomsnittet av de observerte $\hat{\sigma}_i^2$. På en tilsvarende måte setter vi, i tilfellet Jevons indeks og $n_i = 1$,

$$v_i^* = (P_i^{s,t})^2 \{\sum_l I_{n_l>1} n_l v_l / (P_l^{s,t})^2\} / (\sum_l I_{n_l>1}) \tag{20}$$

Under antagelsen om at variasjon i σ_i^2 er tilfeldig fra gruppe til gruppe, så er (19) og (20) omtrent forventningsrette. I tilfellet Dutot indeks og $n_i = 1$, har vi i utgangspunktet

$$v_i^* = (p_{i1}^s)^{-1} \{\sum_l I_{n_l>1} v_l (\sum_j p_{lj}^s)\} / (\sum_l I_{n_l>1})$$

Denne imputeringsmetode er imidlertid noe ustabil der p_{ij}^s enten er veldig stor eller veldig liten. Det er mer robust å erstatte de individe p_{ij}^s med en gjennomsnittlig verdi, som f.eks.

$$\tilde{v}_i^* = (\frac{\sum_l (1 - I_{n_l>1}) p_{l1}^s}{\sum_l (1 - I_{n_l>1})})^{-1} \{\sum_l I_{n_l>1} v_l (\sum_j p_{lj}^s)\} / (\sum_l I_{n_l>1}) \tag{21}$$

Anbefaling Vi anbefaler alternativ III, dvs. direkte imputering av v_i^* ved (19), (20) og (21).

3 Kjedeindeks

`Kjedeindeks` oppstår når en eller flere referanseperioder endres. I tillegg må vi ha med *tidspunktet for endring*, betegnet med d, som ikke nødvendigvis faller sammen med en av referanseperiodene. Formelt definerer vi `kjeding` som en multiplikativ justering av fastvektsindekser. Alle får ikke den samme justeringsfaktor, men hver del-/totalindeks skal få kun én faktor. Forskjellige måter å fastsette denne justeringsfaktor på har forskjellige egenskaper. Spesielt sier vi at kjeding forårsaker et *brudd* dersom den fører til at utviklingen i indeksserie endres, uten at brudd nødvendigvis bør eller kan unngås. F.eks. en Paasche indeks er en kjedeindeks med brudd på hvert tidspunkt ettersom $b = t$ hele tiden endrer seg.

Notasjon Betegn som før med $P_G^{s,t}(b)$ en fastvekts L-indeks med bestemte referanseperioder s, b og $r = s$ for hhv. pris, vekt og indeks. Som regel droppes betegnelsen G for enkelhetsskyld i tilfellet totalindeks. Betegn med $\tilde{P}_G^t$ en kjedeindeks. Som sagt kan ikke alle referanseperioder være entydige i en kjedeindeks. Vi kan ta dem med i notasjonen der det er ønskelig og mulig. F.eks. $\tilde{P}^t(b)$ står for en kjedet serie av totalindekser med ett bestemt sett av basisvekter b.

3.1 Endring i basisperiode for pris

Anta at prisreferanseperioden for en fastvektsindeks $P_G^{s,t}(b)$ endres fra s til s' for $s' > s$. En vanlig grunn til det er at direkte beregning av $P_i^{s,t}$ er blitt umulig pga. fornyelse av varer eller bedrifter. I slike tilfeller har vi som regel $d = s'$. I tillegg virker det naturlig å kreve at kjedeindeksen utvikler seg likt som fastvektsindeksen $P_G^{s,t}(b)$ for $t \le d$, dvs.

$$\tilde{P}_G^{t_2}/\tilde{P}_G^{t_1} = P_G^{s,t_2}(b)/P_G^{s,t_1}(b) \qquad \text{for} \quad t_1 < t_2 \le d \tag{22}$$

Forskjellige valg på kjeding skal kun påvirke utviklingen etter $t = d$, som er lik s' i dette tilfellet. Legg merke til at indeksreferanseperioden r er fortsatt lik s, slik at $\tilde{P}_G^s$ fortsatt må være lik 1.

3.1.1 Elementære indekser

Betrakt følgende kjeding av elementære indekser som tilfredsstiller betingelse (22)

$$\tilde{P}_i^t = \begin{cases} P_i^{s,t} & \text{for } t \le s' \\ P_i^{s,s'} P_i^{s',t} & \text{for } t > s' \end{cases}$$

Medfører kjeding et brudd ved $t = d$? En elementær indeks $P_i^{s,t}$ sies å være *transitiv* dersom

$$P_i^{s,t} = P_i^{s,s'} P_i^{s',t}$$

Både Dutot og Jevons indekser er transitive, men ikke Carli indeksen. Det er klart at kjeding over s' ikke medfører noe brudd i en transitiv elementær indeks.

3.1.2 Aggregering av elementære indekser

L-indeks Følgende kjeding av L-indekser tilfredsstiller betingelse (22) per definisjon,

$$\tilde{P}_G^t(b) = \begin{cases} P_G^{s,t}(b) & \text{for } t \le d = s' \\ P_G^{s,s'}(b) P_G^{s',t}(b) & \text{for } t > d = s' \end{cases} \tag{23}$$

Utviklingen i kjedeindeks etter tidspunktet $d = s'$ er lik fastvektsindeks $P_G^{s',t}(b)$ med den nye prisbasisperiode s'. Dette resulterer i et brudd siden, for $t > s'$ og transitiv elementær indeks,

$$P^{s,t}(b) = \sum_i w_i^b P_i^{s,s'} P_i^{s',t} = P^{s,s'}(b)(\sum_i \tilde{w}_i^{b'} P_i^{s',t}) \neq P^{s,s'}(b) P^{s',t}(b)$$

der $P^{s,t}(b)$ er en L-indeks, og

$$\tilde{w}_i^{b'} = (w_i^b P_i^{s,s'})/(\sum_k w_k^b P_k^{s,s'})$$

er vekten fra basisperiode b prisjustert til periode $b' = b + s' - s$. Det følger at

$$\tilde{P}_G^t = \begin{cases} P_G^{s,t}(b) & \text{for } t \leq d = s' \\ P_G^{s,s'}(b) \tilde{P}_G^{s',t} & \text{for } t > d = s' \quad \text{der} \quad \tilde{P}^{s',t} = \sum_i \tilde{w}_i^{b'} P_i^{s',t} \end{cases} \tag{24}$$

ikke forårsaker noe brudd pga. endring i prisreferanseperioden. Legg merke til at vi i (24) ikke skriver $\tilde{P}^{s',t} = P^{s,t}(b)/P^{s,s'}(b)$ nettopp for å unngå direkte sammenligning av priser for s og t.

P-indeks Kjeding (23) gjelder også for P-indekser. Bare bytt ut P med $\wp$. For (24) har vi

$$\tilde{\wp}^{s',t} = \{\sum_i \tilde{w}_i^{b'} (P_i^{s',t})^{-1}\}^{-1} \quad \text{og} \quad \tilde{w}_i^{b'} = (w_i^b / P_i^{s,s'})/(\sum_k w_k^b / P_k^{s,s'}) \quad \text{og} \quad b' = b - s' + s$$

Laspeyres indeks For en Laspeyres indeks har vi $b = s$, og

$$P^{s,t+1}(s)/P^{s,t}(s) = (\frac{\sum_i q_i^s p_i^{t+1}}{\sum_i q_i^s p_i^s})/(\frac{\sum_i q_i^s p_i^t}{\sum_i q_i^s p_i^s}) = \frac{\sum_i q_i^s p_i^{t+1}}{\sum_i q_i^s p_i^t} = \sum_i \tilde{w}_i^t \frac{p_i^{t+1}}{p_i^t}$$

der q_i og p_i er kvantum og pris på den $i$te vare, og $\tilde{w}_i^t$ er prisjustert fra w_i^s, dvs.

$$\tilde{w}_i^t = (q_i^s p_i^t)/(\sum_k q_k^s p_k^t) = (q_i^s p_i^s I_i^{s,t})/(\sum_k q_k^s p_k^s I_k^{s,t})$$

Laspeyres indeksen kan derfor tolkes som en kjedeindeks (24), der vekten er prisjustert til tidspunkt t ved hver anledning, og den inneholder ingen brudd.

3.1.3 Teoretisk kjedeindeks

Den statistiske modell for elementær indeks $P_i^{s',t}$, for $t > s'$, har den samme form som modellen for $P_i^{s,t}$, for $t \leq s'$, slik at

$$\theta_i^{s,t} = \theta_i^{s,s'} \theta_i^{s',t} \qquad \text{for} \quad t > s'$$

I tilfeller transitive elementære indekser kan man til og med estimere $\theta_i^{s,t}$ under én modell uansett om $t \leq s'$ eller $t > s'$. Endringen i prisbasisperioden har ingen betydning her, men den tillater oss å *oppdatere utvalget*, uansett om varene er trukket med sannsynlighet eller representativt.

For total-/delindeks må man foreta et valg mellom (23) og (24). Forskjellen ligger i $\theta^{s',t}$ for $t > s'$. I tilfellet (23) er den teoretiske indeks gitt ved $\theta^{s',t} = \theta^{s',t}(b)$. Mens vi har, i tilfellet (24),

$$\theta^{s',t} = \sum_i \tilde{\omega}_i^{b'} \theta_i^{s',t} = \frac{\sum_i \omega_i^b \theta_i^{s,s'} \theta_i^{s',t}}{\sum_i \omega_i^b \theta_i^{s,s'}} = (\frac{\sum_i q_i^b p_i^b \theta_i^{s,t}}{\sum_i q_i^b p_i^b})/(\frac{\sum_i q_i^b p_i^b \theta_i^{s,s'}}{\sum_i q_i^b p_i^b}) = \frac{\theta^{s,t}(b)}{\theta^{s,s'}(b)}$$

Hadde en direkte indeks vært en indifferensindeks, ville det ikke har vært noe forskjell om man beregnet den på den ene eller andre måte, siden en indifferensindeks er transitiv per definisjon. Men en direkte indeks er ikke en indifferensindeks. Vi kjenner heller ikke til det eksakte forhold mellom de to. Følgelig kan ikke valget avgjøres på det økonomiske indeksteoretiske grunnlag.

Det man kan undersøke empirisk er vektene, siden forskjellen mellom (23) og (24) ligger i bruken av w_i^b eller $\tilde{w}_i^{b'}$, for $t > d = s'$. Kanhende man mener at vektene burde oppdateres ved tidspunkt $d = s'$, uten at man har tilgang til $w_i^{s'}$ direkte. Anta at omstendigheten er slik at w_i^b er laget som et godt estimat for verdiandelen v_i^b/v^b. Følgelig ønsker man en vekt som er det beste mulige estimat for $v_i^{s'}/v^{s'}$ for $t > s'$, og kan velge mellom w_i^b og $\tilde{w}_i^{b'}$ i tråd med det.

3.2 Endring i basisperiode for vekt

Det er sjelden ønskelig eller mulig å fastholde basisvekten over lang tid. Anta at vektbasisperioden for en fastvektsindeks $P_G^{s,t}(b)$ endres fra b til b' ved tidspunkt d, for $b < b' \leq s < d$. Dette er en situasjon der man oppdaterer vektene uten at det er nødvendig å oppdatere basisprisene. Anta videre betingelse (22), dvs. ingen revisjon for indeksen mellom $t = s$ og $t = d$. Ellers kunne man jo like greit skifte ut $P_G^{s,t}(b)$ med $P_G^{s,t}(b')$, som er mulig siden $b' \leq s$. Legg merke til at endringen i vektbasisperiode ikke påvirker elementære indekser da disse er uavhengige av b. Men brudd kan ikke unngås for høyere aggregeringer.

Alternativ I: Direkte metode Betrakt følgende kjedeindeks som tilfredsstiller (22)

$$\tilde{P}_G^t = \begin{cases} P_G^{s,t}(b) & \text{for } t \leq d \\ P_G^{s,d}(b) P_G^{d,t}(b') & \text{for } t > d \end{cases} \tag{25}$$

Utviklingen i kjedeindeksen etter tidspunktet d er *direkte* gitt ved fastvektsindeksen $P_G^{d,t}(b')$ med vektbasisperiode b' og prisbasisperiode d.

Alternativ II: Indirekte metode Betrakt følgende kjedeindeks med faste basispriser som også tilfredsstiller betingelse (22)

$$\tilde{P}_G^{s,t} = \begin{cases} P_G^{s,t}(b) & \text{for } t \leq d \\ P_G^{s,d}(b) P_G^{s,t}(b')/P_G^{s,d}(b') & \text{for } t > d \end{cases} \tag{26}$$

Utviklingen i kjedeindeksen etter tidspunkt d er *indirekte* gitt ved forholdet mellom to fastvektsindekser, nemlig, $P_G^{s,t}(b')/P_G^{s,d}(b')$. Legg merke til at uttrykket $P_G^{s,t}(b')$ forutsetter at direkte sammenligning mellom p_{ij}^t og p_{ij}^s er mulig for $t > d$. Vi har, i tilfellet L-indekser,

$$\frac{P^{s,t}(b')}{P^{s,d}(b')} = \frac{\sum_i w_i^{b'} P_i^{s,t}}{\sum_i w_i^{b'} P_i^{s,d}} = \sum_i \Big(\frac{w_i^{b'} P_i^{s,d}}{\sum_k w_k^{b'} P_k^{s,d}} \Big) \frac{P_i^{s,t}}{P_i^{s,d}} = \sum_i \tilde{w}_i^{b^*} P_i^{d,t}$$

gitt transitiv elementær indeks, der

$$\tilde{w}_i^{b^*} = (w_i^{b'} P_i^{s,d}) / (\sum_k w_k^{b'} P_k^{s,d}) \qquad \text{for} \qquad b^* = b' + d - s$$

er vekten i basisperiode b' prisjustert til b^*. Mens vi har, i tilfellet P-indekser,

$$\frac{\wp^{s,t}(b')}{\wp^{s,d}(b')} = \frac{\sum_i w_i^{b'}/P_i^{s,d}}{\sum_i w_i^{b'}/P_i^{s,t}} = \sum_i \Big(\frac{w_i^{b'}/P_i^{s,t}}{\sum_k w_k^{b'}/P_k^{s,t}} \Big) \frac{P_i^{s,t}}{P_i^{s,d}} = \sum_i \tilde{w}_i^{b^*} P_i^{d,t}$$

gitt transitiv elementær indeks, der $\tilde{w}_i^{b^*}$ er prisjustert fra $w_i^{b'}$ ved

$$\tilde{w}_i^{b^*} = (w_i^{b'}/P_i^{s,t})/(\sum_k w_k^{b'}/P_k^{s,t}) \qquad \text{for} \qquad b^* = b' - t + s$$

Paasche indeks I Paasche indeksen skiftes vektbasisperioden $b = t$ hele tiden. Den er derfor kjedet sammen og har brudd ved hvert tidspunkt. Vi har

$$\frac{P^{s,t+1}(t+1)}{P^{s,t}(t)} = (\frac{\sum_i q_i^{t+1} p_i^{t+1}}{\sum_i q_i^{t+1} p_i^s})/(\frac{\sum_i q_i^t p_i^t}{\sum_i q_i^t p_i^s}) = (\frac{\sum_i q_i^{t+1} p_i^{t+1}}{\sum_i q_i^t p_i^t})/(\frac{\sum_i q_i^{t+1} p_i^s}{\sum_i q_i^t p_i^s}) = \frac{V^{t,t+1}}{Q^{t,t+1}(s)}$$

der q_i og p_i er kvantum og pris på den $i$te vare, og $V^{t,t+1}$ er verdiindeksen mellom t og $t+1$, og $Q^{t,t+1}(s)$ er Laspeyres volumindeksen mellom t og $t+1$. Mao. den velkjente dekomponering av verdiindeks mellom t og $t+1$ som produkt av volum- og prisindekser.

Teoretisk kjedeindeks Det er ingen endringer i den underliggende statistiske modell for en elementær indeks. Den teoretiske indeks $\theta^{d,t}$ for $t > d$ er gitt ved $\theta^{d,t}(b')$ i tilfellet (25), og den er gitt ved $\theta^{s,t}(b')/\theta^{s,d}(b')$ i tilfellet (26). Heller ikke her kan valget mellom de to generelt avgjøres mht. økonomiske indeksteorien. Men det kan være så at man ønsker at vekten skal være et godt estimat for verdiandelen v_i^d for $t > d$, og kan velge mellom $w_i^{b'}$ og $\tilde{w}_i^{b*}$ deretter.

3.3 Endring i basisperioder for vekt og pris

Anta at vektbasisperioden for en fastvektsindeks $P_G^{s,t}(b)$ endres fra b til b' ved tidspunkt d, samtidig som prisbasisperioden endres fra s til s', der

$$b < b' \quad \text{og} \quad s < s' \quad \text{og} \quad b < s \quad \text{og} \quad b' < s'$$

Dette er f.eks. situasjonen i KPI, der både basisvektene og -prisene oppdateres årlig, og vektene er basert på forbruksundersøkelser fra tidligere årganger. I tillegg kan man anta at $d = s'$, da det er ingen grunn til å vente med kjeding til et senere tidspunkt. Anta videre betingelse (22), dvs. man ønsker ikke å revidere indeksen mellom $t = s$ og $t = d = s'$. Da er (25) det naturlige valg.

3.4 Endring i indeksreferanseperiode: Renormering

Renormering skjer utelukkende i forbindelse med presentasjon av prisindekser.

3.4.1 Renormering via separat skalering

Renormering av fastvektsindeks Anta først at indeksreferanseperioden for en fastvektsindeks $P_G^{s,t}(b)$ endres fra $r = s$ til $r' \neq s$. Etterpå har vi $\tilde{P}_G^s \neq 1$ men $\tilde{P}_G^{r'} = 1$. Betrakt følgende skalering (*separat* for forskjellig G) som tilfredsstiller betingelse (22)

$$\tilde{P}_G^t = P_G^{s,t}(b)/P_G^{s,r'}(b) \tag{27}$$

Utviklingen i den renormerte serie $\tilde{P}_G^t$ er identisk med den gamle serie $P_G^{s,t}(b)$ for alle t og d, dvs. ingen brudd. Imidlertid er det slik at de renormerte elementære indekser ikke kan aggregeres opp til et høyere nivå vha. w_i^b, med mindre $P_i^{s,r'}$ er like for alle i.

Aggregering av elementære indekser Vi har i tilfellet L-indekser

$$\tilde{P}^t = \frac{P^{s,t}(b)}{P^{s,r'}(b)} = \frac{\sum_i w_i^b P_i^{s,t}}{\sum_i w_i^b P_i^{s,r'}} = \sum_i \frac{w_i^b P_i^{s,r'}}{\sum_k w_k^b P_k^{s,r'}} \frac{P_i^{s,t}}{P_i^{s,r'}} = \sum_i \frac{w_i^b P_i^{s,r'}}{\sum_k w_k^b P_k^{s,r'}} \tilde{P}_i^t \neq \sum_i w_i^b \tilde{P}_i^t$$

Dersom elementære indeksen er transitiv, har vi

$$\tilde{P}_G^t = \sum_i \tilde{w}_{i(G)}^{b'} P_i^{r',t} \qquad \text{der} \quad \tilde{w}_{i(G)}^{b'} = \frac{w_i^b P_i^{s,r'}}{\sum_{k \in G} w_k^b P_k^{s,r'}} \quad \text{og} \quad b' = b + r' - s \tag{28}$$

Dvs. renormering av en L-indeks ved (27) er det samme som å beregne den ved (28) med prisjusterte vekter $\tilde{w}_{i(G)}^{b'}$ og prisbasisperiode $s = r'$. I tilfellet P-indekser, har vi

$$\wp^t = \frac{\{\sum_i w_i^b (P_i^{s,t})^{-1}\}^{-1}}{\{\sum_i w_i^b (P_i^{s,r'})^{-1}\}^{-1}} = \frac{\sum_i w_i^b / P_i^{s,r'}}{\sum_i w_i^b / P_i^{s,t}} = \sum_i \frac{w_i^b / P_i^{s,t}}{\sum_k w_k^b / P_k^{s,t}} \frac{P_i^{s,t}}{P_i^{s,r'}}$$

slik at renormering av en P-indeks ved (27) er det samme som å beregne følgende L-indeks

$$\tilde{P}_G^t = \sum_i \tilde{w}_{i(G)}^{b'} P_i^{r',t} \qquad \text{der} \quad \tilde{w}_{i(G)}^{b'} = \frac{w_i^b / P_i^{s,t}}{\sum_{k \in G} w_k^b / P_k^{s,t}} \quad \text{og} \quad b' = b - t + s \tag{29}$$

Diskusjon Betrtakt følgende renormering gitt transitiv elementær indeks

$$\tilde{P}_G^t = P_G^{r',t}(b) = \sum_{i \in G} w_{i(G)}^b P_i^{r',t} = \sum_{i \in G} w_{i(G)}^b (P_i^{s,t} / P_i^{s,r'}) \tag{30}$$

Etterpå er $\tilde{P}^{t'} / \tilde{P}^t$ gitt ved forholdet mellom to fastvektsindekser for t' og t med samme referanseperiodene $(s = r', b, r = r')$. Derfor oppfyller ikke renormering (30) betingelse (22), og revisjon er nødvendig. Det er klart at (30) fører til et brudd ved tidspunkt r' siden, for $t_1 < r' < t_2$,

$$\frac{\tilde{P}_G^{t_2}}{\tilde{P}_G^{t_1}} = \frac{P^{r',t_2}(b)}{P^{r',t_1}(b)} = (\sum_i w_i^b P_i^{r',t_2}) \{\sum_i w_i^b (P_i^{t_1,r'})^{-1}\}^{-1}$$

dvs. et produkt av $\wp^{t_1,r'}$ og P^{r',t_2}, der vi antar tidsombyttbar elementær indeks $P_i^{r',t_1} = 1/P_i^{t_1,r'}$.

Renormering for kjedeindeks Separat skalering (27) er det naturlige valg i tilfellet renormering i tillegg til endringer i pris- eller/og vektbasisperiode, dvs. kjedeindeks. Vi har

$$\tilde{P}_G^t(r') = \tilde{P}_G^t(r) / \tilde{P}_G^{r'}(r) \tag{31}$$

Altervativet (30) fører til omfattende revisjon; det kan også være direkte umulig i en del situasjoner. Når det gjelder kjedingsrekkefølge mht. referansperiodene, anbefaler vi at man

- først kjede over enderinger i basisvektene og/eller -prisene
- deretter foreta renormering av den kjedete indeksserie.

Dette er mer oversiktig enn den omvendte rekkefølge spesielt i tilfellet $r' \neq s'$.

3.4.2 Renormering over flere perioder

Det hender ved renormering at man ønsker å sette den gjennomsnittlige prisindeks over flere perioder til 1. F.eks. man kan sette den gjennomsnittlige prisindeks for et helt år til 1 for en månedelig indeks. La R betegne et sett av m perioder, slik at etter renormering skal vi ha

$$m^{-1} \sum_{t \in R} \tilde{P}^t = 1$$

Betrakt følgende separat skalering av en fastvektsindeks $P_G^{s,t}(b)$

$$\tilde{P}_G^t = \frac{P_G^{s,t}(b)}{m^{-1} \sum_{t' \in G} P_G^{s,t'}(b)} \tag{32}$$

som i likhet med (27) bevarer utviklingen i den opprinnelige indeks. For en L-indeks har vi

$$\tilde{P}^t = \frac{\sum_i w_i^b P_i^{s,t}}{m^{-1} \sum_{t' \in R} (\sum_i w_i^b P_i^{s,t'})} = \frac{\sum_i w_i^b P_i^{s,t}}{\sum_i w_i^b (m^{-1} \sum_{t' \in R} P_i^{s,t'})} = \frac{\sum_i w_i^b P_i^{s,t}}{\sum_i w_i^b \bar{P}_i^{s,R}}$$

der $\bar{P}_i^{s,R} = m^{-1} \sum_{t' \in R} P_i^{s,t'}$. Mao. renormering ved (32) er det samme som å beregne L-indeks

$$\tilde{P}_G^t = \sum_i \tilde{w}_{i(G)} \tilde{P}_i^t \qquad \text{der} \quad \tilde{w}_{i(G)} = \frac{w_i^b \bar{P}_i^{s,R}}{\sum_{k \in G} w_k^b \bar{P}_k^{s,R}} \quad \text{og} \quad \tilde{P}_i^t = \frac{P_i^{s,t}}{\bar{P}_i^{s,R}} \tag{33}$$

Legg merke til at $\tilde{P}_i^t$ i (33) står i samsvar med (32). For en P-indeks har vi

$$\tilde{\wp}^t = \frac{\frac{1}{\sum_i w_i^b / P_i^{s,t}}}{\frac{1}{m} \sum_{t' \in R} \frac{1}{\sum_i w_i^b / P_i^{s,t'}}} = \frac{1}{\frac{1}{m} \sum_{t' \in R} \frac{\sum_k w_k^b / P_k^{s,t}}{\sum_i w_i^b / P_i^{s,t'}}} = \frac{1}{\frac{1}{m} \sum_{t' \in R} (\sum_i \frac{w_i^b / P_i^{s,t}}{\sum_k w_k^b / P_k^{s,t}} \frac{P_i^{s,t}}{P_i^{s,t'}})^{-1}}$$

slik at renormering ved (32) er det samme som å beregne

$$\tilde{\wp}_G^t = \{m^{-1} \sum_{t' \in R} (\sum_i \tilde{w}_{i(G)}^{b'} / P_i^{t,t'})^{-1}\}^{-1} = \{m^{-1} \sum_{t' \in R} \tilde{\wp}^{t,t'}\}^{-1} \tag{34}$$

dvs. invers til gjennomsnittet av P-indekser (t, t') for $t' \in R$ med prisjusterte vekten

$$\tilde{w}_{i(G)}^{b'} = \frac{w_i / P_i^{s,t}}{\sum_{k \in G} w_k / P_k^{s,t}} \qquad \text{der} \qquad b' = b - t + s$$

3.5 Eksempler

I praksis kan kjeding tilsynelatende forekomme på andre måter enn de som er presentert ovenfor, og sammenhengen blir først klar etter en nærmere analyse. La oss se på to eksempeler.

3.5.1 Kjeding for KPI

Formalisering Konsumprisindeksen (KPI) er en månedelig kjedeindeks. Basisvekten og -prisen skiftes en gang i året, der vektene er basert på forbruksundersøkelser fra tidligere årganger. Formelt betrakter vi en situasjon der referanseperioder (b_α, s_α) endres til $(b_{\alpha+1}, s_{\alpha+1})$ ved tidspunktet $d = s_{\alpha+1}$, og $b_\alpha < s_\alpha$ for alle α som går fra 1 til det uendelige. Vi trenger bare å utrede kjeding mellom to fastvektsindekser; resten følger induktivt deretter.

Kjeding uten renormering Betegn to fastvektsindekser $P_G^{s_0,t}(b_0)$ for $t \in (s_0, s_1]$ og $P_G^{s_1,t}(b_1)$ for $t \in (s_1, s_2]$. Mht. betingelse (22) bør kjeding for KPI være slik at

$$\tilde{P}^{t_2}/\tilde{P}^{t_1} = \begin{cases} P^{s_0,t_2}(b_0)/P^{s_0,t_1}(b_0) & \text{for } t_1 < t_2 \leq d = s_1 \\ P^{s_1,t_2}(b_1)/P^{s_1,t_1}(b_1) & \text{for } d = s_1 \leq t_1 < t_2 \end{cases}$$

Anta i tillegg følgende *betingelse for kumulative endringer* i tilfellet $t_1 < d = s_1 < t_2$

$$\frac{\tilde{P}^{t_2}}{\tilde{P}^{t_1}} = \frac{P^{s_0,s_1}(b_0)}{P^{s_0,t_1}(b_0)} \frac{P^{s_1,t_2}(b_1)}{P^{s_1,s_1}(b_1)} = \frac{P^{s_0,s_1}(b_0)}{P^{s_0,t_1}(b_0)} P^{s_1,t_2}(b_1) \tag{35}$$

dvs. et produkt av endringen fra t_1 til $d = s_1$ ifølge den gamle serie og endringen fra $d = s_1$ til t_2 ifølge den nye serie. Legg merke til at den siste lingning på høyre siden følger av at $P^{s_1,s_1}(b_1) = 1$. Det er nå lett å verifisere at metode (25) tilfredsstiller alle betingelsene ovenfor.

Kjeding med renormering Anta at indeksreferanseperioden endres fra r_0 til $r_1 = s_1 = d$ samtidig som (b_0, s_0) endres til (b_1, s_1). En vanlig praksis er å renormere gamle serien ved

$$\tilde{P}_G^t = P_G^{s_0,t}(b_0)/P_G^{s_0,d}(b_0) \qquad \text{for} \quad t \leq d = s_1$$

slik at $\tilde{P}_G^d = P_G^{s_1,d}(b_1) = 1$. Den renormerte gamle serie er dermed kjedet sammen med den nye serie ved tidspunkt d. Praksisen er akkurat den samme som kjeding ved (25) og renormering ved (27) deretter. Rekkefølgen i skifting av (b, s) og r er uvesentlig i dette tilfelle.

3.5.2 Kjeding for PIF i 2000

Formalisering Prisindeksen for førstegangsomsetning (PIF) innenlands har vært en månedelig fastvektsindeks fram til desember 2000. I 2000 innførte man bl.a. en ny aggregeringsstruktur og dertil nye basisvekter og -priser. Den nye fastvektsindeks ble beregnet gjennom hele året 2000, slik at ved desember 2000 hadde man to parallelle indeksserier over 12 terminer. Til tross for den noe spesielle omstendighet rundt PIF i 2000 er oppdatering av referanseperioder (b, s) formelt sett helt identisk som i tilfellet KPI ovenfor, der referanseperiodene (b_α, s_α) endres til $(b_{\alpha+1}, s_{\alpha+1})$ ved tidspunkt $d = s_{\alpha+1}$, og $b_\alpha < s_\alpha$ for alle α som går fra 1 til det uendelige. Det spesielle er at man nå legger merke til at det er mulig å beregne parallelt to indeksserier som f.eks. $P^{s_0,t}(b_0)$ og $P^{b_1,t}(b_1)$ for $t \in R = \{b_1 + 1, ..., s_1\}$. Situasjonen er ganske allminnelig så lenge de samme aggregeringene gjelder både for (b_0, s_0) og (b_1, s_1).

Kjeding over tidspunkt s_1 I utgangspunktet er det mulig å kjede over et vilkårlig tidspunkt fra R, da all informasjon som trengs er tilgjengelig. Men i praksis er det kun kjeding over $s_1 = d$ som kan godtas mht. betingelse (22), dvs. ingen revisjon fram til tidspunkt d.

Kjeding med renormering over R Dette er tilfellet PIF i 2000, da man ønsker å renormere kjedeindeksen slik at den gjennomsnittlige indeks over hele år 2000 er lik 1 etterpå. Betrakt følgende framgangsmåte, der man renormerer den gamle serie ved

$$\tilde{P}_{gml}^t = P_{gml}^{s_0,t}(b_0)/\bar{P}_{gml}^{s_0,R}(b_0) \qquad \text{for} \quad t \leq d = s_1 \quad \text{og} \quad \bar{P}_{gml}^{s_0,R}(b_0) = |R|^{-1} \sum_{t'} P_{gml}^{s_0,t'}(b_0)$$

og $\bar{P}^{s_0,R}_{gml}(b_0)$ er gjennomsnittet til gamle indeksene over R med referanseperiodene $(b = b_0, s = s_0)$. Samtidig renormerer man den nye serie ved

$$\tilde{P}^t_{ny} = P^{b_1,t}_{ny}(b_1)/\bar{P}^{b_1,R}_{ny}(b_1) \qquad \text{for} \quad t \geq b_1 \quad \text{og} \quad \bar{P}^{b_1,R}_{ny}(b_1) = |R|^{-1}\sum_{t'} P^{b_1,t'}_{ny}(b_1)$$

der $\bar{P}^{b_1,R}_{ny}(b_1)$ er gjennomsnittet til nye indeksene over R med referanseperiodene $(b = b_1, s = b_1)$. På denne måte er både den gamle og den nye serie renormert til 1 over R. Den gamle serie fram til $t \leq d = s_1$ er dermed kjedet sammen med den nye serie fra $t > d = s_1$ ved tidspunktet $d = s_1$. Det følger at kjedeindeksen utvikler seg likt den gamle serie før tidspunktet d, og den utvikler seg likt den nye serie etter tidspunktet d. For $t_1 < d = s_1 < t_2$, har vi derimot

$$\frac{\tilde{P}^{t_1}}{\tilde{P}^{t_2}} = \frac{P^{s_0,s_1}(b_0)}{P^{s_0,t_1}(b_0)} \frac{P^{b_1,t_2}(b_1)}{\bar{P}^{b_1,R}(b_1)}$$

som ikke tilfredsstiller betingelse (35) for kumulative endringer. Derfor bør man først kjede over endringen fra (b_0, s_0) til (b_1, s_1) ved (25), og deretter foreta renormering ved (32), dvs. $\tilde{P}^t_G/\bar{\tilde{P}}^R_G$, i tråd med anbefalingen som følger under (31).

4 Hedonisk indeks

Byggesteinen i en fastvektsindeks er prisparet av samme varen over tid (eller en annen dimensjon). Men som det innledningsvis har blitt påpekt, to mengder varer (eller tjenester) som sammenfattes i en prisindeks består som regel ikke av samme individe varene. Hedoniske[2] metoder gir oss en mulighet til å utvide teorien til en mer generell situasjon.

Notasjon Vi kan begrense problemet til beregning av en elementær indeks. Aggregering av elementære indekser påvirkes ikke av diskusjonen i dette avsnitt: endring i varemengden medfølger ikke endring i vekten. Det siste betraktes som et problem for kjeding. I notasjonen kan vi droppe gruppeindikator i. F.eks. anta at det observeres n varer, betegnet med $j = 1, ..., n$. Basert på disse beregner vi $P^{s,t}$, eller ganske enkelt P, ved en elementær indeksformel uten bruk av vekt.

4.1 Tradisjonelle ikke-hedoniske metoder

Anta at det observeres n varer på tidspunkt s. Anta først at ved tidspunkt $s + 1$ er den $n$te vare tatt ut av utvalget, kanskje til og med populasjonen.

Gjør-ingenting metode I den enkleste situasjon trenger man ikke å gjøre noe annet enn å beregne $P^{s,s+1}$ kun basert på $j = 1, ..., n - 1$. Det antas da at vareutskifting ikke endrer på den underliggende statistiske modell, og utvalget er like representativt, uansett om den $n$te vare fremdeles finnes ute i populasjonen eller ikke.

Overlappende linke metode Anta at en ny vare er tatt inn i utvalget, betegnet med n'. I tilfellet den er ny i populasjonen fra og med tidspunktet $s + 1$, skal den være en del av beregningsgrunnlaget for $P^{s+1,s+2}$, osv. Men den kan ikke være med i beregningen av $P^{s,s+1}$ under en av de 3 statistiske modeller vi nå har, da disse forutsetter at varen finnes i populasjonen på begge tidspunkter. I tilfellet n' allerede finnes i populasjonen ved tidspunkt s, kan man beregne $P^{s,s+1}$ basert på $j = 1, ..., n - 1, n'$, dersom $p^s_{n'}$ er tilgjengelig så vel som $p^{s+1}_{n'}$. Dette er den *overlappende linke metode*. Det antas at det nye utvalg $\{1, ..., n - 1, n'\}$ er representativt som det gamle utvalg $\{1, ..., n - 1, n\}$ var. Men i praksis er det ikke alltid mulig å få tak i $p^s_{n'}$ retrospektivt, hvis vare n' første gang inkluderes i utvalget ved et senere tidspunkt.

Direkte imputeringsmetode Anta at n' allerede finnes i populasjonen ved tidspunkt s, men blir først med i utvalget ved tidspunkt $s + 1$, og $p^s_{n'}$ er utilgjengelig. Man kan *imputere* for $p^s_{n'}$, betegnet med $p^{*s}_{n'}$. Med *direkte imputeringsmetoden* setter man

$$p^{*s}_{n'} = p^s_n \qquad \rightarrow \qquad I^{*s,s+1}_{n'} = p^{s+1}_{n'}/p^s_n$$

noe som medfører en skjevhet i den imputerte indeks $P^{*s,s+1}$ med mindre $p^s_n = p^s_{n'}$.

Imputer-ingen-endring metode I denne metode setter man

$$p^{*s}_{n'} = p^{s+1}_{n'} \qquad \rightarrow \qquad I^{*s,s+1}_{n'} = 1$$

som igjen medfører en skjevhet i P^* med mindre $E(I^{s,s+1}_{n'}) = 1$.

[2]Ifølge Andrew Court, forslo Andre Sachs ordet "hedonic" basert på antagelsen om at hedoniske indekser måler "the potential contribution ... to the welfare and happiness of its purchaser and the community".

Imputer-gruppe-endring metode Først velger man en mengde $B \subseteq \{1, ..., n-1\}$, og beregner $P_B^{s,s+1}$ basert på varer som tilhører B. Deretter setter man

$$p_{n'}^{*s} = p_{n'}^{s+1}/P_B^{s,s+1} \qquad \rightarrow \qquad I_{n'}^{*s,s+1} = P_B^{s,s+1}$$

Mao. man antar at vare n' har den samme prisutvikling som varene i mengde B. Antagelsen er *ad hoc*, da det ikke tillates å innføre flere statistiske modeller innen en elementær gruppe under beregningsopplegget for fastvektsindekser. I tilfellet $B = \{1, ..., n-1\}$ er den imputerte $P^{*s,s+1}$ lik $P^{s,s+1}$ ved gjør-ingenting metoden. Imputering oppnår ingenting i tillegg, sett bort fra en mulig kilde til forvirring i variansestimering. Det blir feil hvis man behandler $p_{n'}^{*s}$ på lik linje med observerte $(p_1^s, ..., p_{n-1}^s)$. Da er det bedre å bruke gjør-ingenting metoden.

4.2 Karakteristikk indeks

4.2.1 Hedonisk funksjon

En `hedonisk funksjon` er en *statistisk modell* for priser på forskjellige varianter av en vare gitt deres *karakteristikk*. Datamaskin er et vanlig eksempel på en slik vare med mange varianter, hvis viktigste karakteristikk inkluderer CPU hastighet, minne, osv. Tidligere studier av hedoniske metoder tok ofte for seg priser på biler. Betegn med p_j prisen til varevariant j ved et bestemt tidspunkt. Betegn med $x_{j1}, ..., x_{jK}$ et sett av K karakteristikker.

Lineær hedonisk funksjon En `lineær` hedonisk funksjon er gitt ved

$$p_j = \beta_0 + \beta_1 x_{j1} + \cdots + \beta_K x_{jK} + \epsilon_j$$

Residualen ϵ_j er en stokastisk variabel med forventning null. Vanligvis antar man at $V(\epsilon_j) = \sigma^2$ er en konstant, men det er mulig å anta andre variansstrukturer. Regresjonskoeffisientene $\beta_1, ..., \beta_K$ tolkes som den *teoretiske* pris til karakteristikk $x_{j1}, ..., x_{jK}$, dvs. hvor mye prisen p_j varierer med en bestemt karakteristikk når alle andre karateristikkere holdes faste. Koeffisienten β_0 kan tolkes som en ikke-navngitt karakteristikk hvis kvantum alltid er lik 1 for én vare. Siden en slik karakteristikk ikke kan endres, har β_0 ingenting med prisvariasjonen å gjøre, men er nødvendig for å kunne bestemme prisnivået.

Log-log hedonisk funksjon En tilsvarende `log-log` hedonisk funksjon er gitt ved

$$\log p_j = \beta_0 + \beta_1 \log x_{j1} + \cdots + \beta_K \log x_{jK} + \epsilon_j$$

der både avhengige og uavhengige variablene transformeres til logaritmiske skala. Vi har

$$\frac{\partial \log p_j}{\partial x_{j1}} = \frac{1}{p_j}\frac{\partial p_j}{\partial x_{j1}} = \beta_1 \frac{1}{x_{j1}} \qquad \Leftrightarrow \qquad \frac{\partial p_j}{p_j} = \beta_1 \frac{\partial x_{j1}}{x_{j1}}$$

Dvs. β_1 angir den relative prisendring mht. relative endringen i x_{j1}, også kjent som *elastisiteten*. Lignende gjelder for $\beta_2, ..., \beta_K$ mht. $x_{j2}, ..., x_{jK}$. Tolkningen av β_0 er det samme som før.

Semi-log hedonisk funksjon En `semi-log` hedonisk funksjon er gitt ved

$$\log p_j = \beta_0 + \beta_1 x_{j1} + \cdots + \beta_K x_{jK} + \epsilon_j$$

der kun prisen er tranformert til log-skala. Koeffisienten β_k angir den relative prisendring mht. endringen i x_{jk}, for $k = 1, ..., K$. Koeffisienten β_0 har den samme tolkning som før.

4.2.2 Vare og karakteristikk

En hedonisk funksjon ‘deler opp’ en vare i et sett av karakteristikker. Det gjør det mulig å sammenligne priser på varer som ikke er identiske, men kan karakteriseres på samme måten. Dermed frigjøring fra antagelsen om identiske varer. F. eks. kan forskjellen mellom varer n og n' nå kvantifiseres som forskjellen mellom $(x_{n1}, ..., x_{nK})$ og $(x_{n'1}, ..., x_{n'K})$.

Men det er ikke opplagt hvordan en vare skal karakteriseres. Dette er ikke bare et empirisk spørmål. Man kan tilsynlatende skille mellom karakteristikk som er fysisk og målbar, som f.eks. minne i antallet MB, og karakteristikk som ikke er fysisk målbar, som f.eks. betydningen av merkevare. Som regel er slike ikke-målbare karakteristikker med på å bestemme prisen, eller prisvariasjonen, og de er i så måte relevante. Det å sette kvantum til 1 på summen av alle slike karakteristikker er derfor generelt en misspesifikasjon av den hedoniske funksjon.

Under hedoniske framgangsmåten betraktes en vare som sammensatte karakteristikker. Man går ikke direkte løs på spørsmålet om hvordan et sett av karakteristikker er satt sammen til en vare, men konsentrerer seg om hvordan prisen (eller prisvariasjonen) på en vare avgjøres av karakteristikkene. Er den hedoniske funksjon i form av en lineær prediktor en tilstrekkelig måte å fange dette opp på? Antageligvis ikke. Ta f.eks. en datamaskin med to karakteristikker, nemlig, CPU-hastighet og minne. Betydning av en økning i CPU-hastigheten varierer etter hvor stort minne maskinen har. Omvendt gjelder det økning i minne gitt CPU-hastigheten. I en lineær prediktor har en regresjonskoeffisient samme effekten uansett hvilket nivå andre uavhengige variabler ligger på. Dette er en sterk antagelse sett i lyset av hvordan datamaskinskapasiteten endres med dens karaketeristikk. Problemet kan ikke håndteres f.eks. med å innføre et interaksjonsledd i prediktoren som i en vanlig regresjonsanalsye. For hva er et produkt (eller annen funksjon) av CPU-hastighet og minne? Er det en ny karakteristikk? Er den målbar? Kort sagt er dette en annen form for misspesifikasjon av hedoniske funksjonen.

På en måte kan de ovennevnte problemer betraktes som nye sider i problemet om målenhet.

4.2.3 Indeksformel

Basert på lineære hedoniske funskjoner Anta følgende lineære hedoniske funksjoner for statistikkperiode t og prisbasisperiode s, nemlig,

$$p_j^t = (x_j^t)^T \beta^t + \epsilon_j^t \qquad \text{og} \qquad p_j^s = (x_j^s)^T \beta^s + \epsilon_j^s$$

der T betegner matrise-/vektor transponering, og $x_j^t = (x_{j0}^t, x_{j1}^t, ..., x_{jK}^t)^T$ der $x_{j0} \equiv 1$, og $\beta^t = (\beta_0^t, \beta_1^t, ..., \beta_K^t)^T$, og lignende for x_j^s og β^s. Modellene har samme form men hver sine tidsbestemte regresjonskoeffisienter. Betegn med $\hat{\beta}^s$ estimatene for β^s, og tilsvarende for tidspunkt t. La a_k være kvantum til den $k$te karakteristikk, den tilsvarende `karakteristikk prisindeks` beregnes ved en *Lowe* indeksformel.

$$K^{s,t} = (\sum_{k=0}^{K} a_k \hat{\beta}_k^t)/(\sum_{k=0}^{K} a_k \hat{\beta}_k^s) \tag{36}$$

Lowe indeksen blir en Laspeyres indeks med kvanta fra prisbasisperioden, mens den blir en Paasche indeks hvis kvanta er hentet fra statistikkperioden. Den tilsvarende teoretiske indeks er gitt ved

$$\theta^{s,t} = (\sum_{k=0}^{K} a_k \beta_k^t)/(\sum_{k=0}^{K} a_k \beta_k^s)$$

Karakteristikk indeks (36) legger et stort krav på datagrunnlaget. Man trenger å vite hvor mange forskjellige varevarianter som omsettes i markedet. Indeksen har f.eks. blitt anvendt for datamaskiner, der en slik markedsoversikt til en viss grad er mulig. Men å gjøre noe lignende er vanskelig for de fleste forbruksvarer (eller tjenester).

Indeksteori Den funksjonelle tilnærming til prisindeks bygger på en teori om nytte. Hedoniske framgangsmåten krever en eiendommelig nytteteori. Istedenfor datamaskin er det nå karakteristikkene som CPU-hastighet og minne som er ettertraktet. Men er det i det hele tatt mulig å snakke om nytten av enkelte karakteristikker, uten at de er satt sammen i en bestemt vare? Legg også merke til at teorien om indifferensindeks ble innført før hedoniske metodene. Den forutsetter ikke identiske mengder av varer, slik en fastvektsindeks gjør. Varianter av en vare med forskjellige karakteristikker kan ganske enkelt betraktes som forskjellige varer med hver sin pris, så lenge man klarer å velge mellom dem. Så spørsmålet er om det er nødvendig å være klar over den eksakte oppdeling av en vare i dens karakteristikk, for å kunne bedømme dens nytteverdi? Oppdelingen er i prinsippet unødvendig hvis svaret er nei. Hvis svaret er ja, blir man nødt til å kartlegge hvordan et sett av karakteristikker er satt sammen til en vare, i tillegg til hvordan man velger mellom forskjellige varer med bestemte karakteristikker. Er det realistisk?

Basert på semi- eller log-log hedonisk funksjon Man kan også bruke en semi- eller log-log hedonisk funksjon. Imidlertid kan ikke en regresjonkoeffisient til en bestemt karakteristikk tolkes som dens pris direkte. Man velger da først én *standard* vare med bestemte kvanta for alle karakteristikkene, betegnet med a_k for $k = 0, ..., K$, som enten representerer en bestemt smak eller et gjennomsnitt i populasjonen. Deretter beregnes den karakteristikk indeks ved

$$K^{s,t} = \exp(\sum_{k=0}^{K} a_k \hat{\beta}_k^t) / \exp(\sum_{k=0}^{K} a_k \hat{\beta}_k^s)$$

i tilfellet semi-log hedonisk funksjon. I tilfellet log-log hedonisk fuknsjon har vi

$$K^{s,t} = \exp(\sum_{k=0}^{K} \hat{\beta}_k^t \log a_k) / \exp(\sum_{k=0}^{K} \hat{\beta}_k^s \log a_k)$$

4.2.4 Variansestimering

Vi kan skrive (36) om på en måte som ligner en fastvektsindeks, dvs.

$$K^{s,t} = \sum_{k=0}^{K} \frac{a_k \hat{\beta}_k^s}{\sum_{l=0}^{K} a_l \hat{\beta}_l^s} \frac{\hat{\beta}_k^t}{\hat{\beta}_k^s} = \sum_k c_k I_k^{s,t} \qquad \text{for} \quad c_k = \frac{a_k \hat{\beta}_k^s}{\sum_{l=0}^{K} a_l \hat{\beta}_l^s} \quad \text{og} \quad I_k^{s,t} = \frac{\hat{\beta}_k^t}{\hat{\beta}_k^s}$$

For å gjøre variansen sammenlignbar med variansen for en fastvektsindeks, betrakter vi $(c_k, \hat{\beta}_k^s)$ som konstante, for $k = 0, ..., K$, slik at variasjonen i $K^{s,t}$ skyldes utelukkende variasjoner i $\hat{\beta}_0^t, ..., \hat{\beta}_K^t$. Dette står forsåvidt i samsvar med realiteten der $\hat{\beta}^s$ estimeres en gang for prisbasisperioden og holdes deretter fast for flere statistikkperioder. I vektorform har vi

$$K^{s,t} = \eta^T \hat{\beta}^t \qquad \text{der} \quad \eta^T = (\frac{c_0}{\hat{\beta}_0^s}, ..., \frac{c_K}{\hat{\beta}_K^s})$$

Betegn med X designmatrisen under den aktuelle karakteristikk hedoniske funksjon, hvis rader svarer til prisobservasjonene ved tidspunkt t og kolonner svarer til karakteristikkene. Betegn med

y kolonnevektoren av avhengige variabler, som er en funksjon av alle observerte priser. F.eks. i tilfellet lineær funksjon har vi $y^T = (p_1^t, ..., p_n^t)$, mens i tilfellet semi-log funksjon har vi $y^T = (\log p_1^t, ..., \log p_n^t)$, der n er antallet priser. Da er $\hat{\beta}^t$ gitt ved

$$\hat{\beta}^t = (X^T X)^{-1}(X^T y) \qquad \Rightarrow \qquad K^{s,t} = \phi^T y \qquad \text{der} \quad \phi^T = \eta^T (X^T X)^{-1} X^T$$

Dermed er variansen til $K^{s,t}$ gitt ved

$$V(K^{s,t}) = \phi^T Cov(y, y)\phi = \phi^T \text{Diag}(\sigma_i^2)\phi$$

der Diag betegner en diagonal matrise og σ_i^2 er variansen til y_i, siden $Cov(y_i, y_j) = 0$ for $i \neq j$. En tilnærmet forventningsrett sandwich variansestmator er gitt ved

$$v(K^{s,t}) = \phi^T \text{Diag}(\frac{e_i^2}{1 - h_i})\phi \qquad \text{der} \quad e = y - X\hat{\beta}^t \quad \text{og} \quad H = X(X^T X)^{-1} X^T$$

og h_i er det $i$te element på diagonalen til matrise H.

4.3 Dummy-variabel indeks

4.3.1 Hedonisk funksjon og indeksformel

Betrakt to påfølgende perioder s og $s+1$, som hhv. er prisbasis- og statistikkperioden. Betegn med x_j^t en kolonnevektor med alle K karakteristikkene til vare j ved tidspunkt t, dvs. $x_j^t = (x_{j0}, x_{j1}, ..., x_{jK})^T$, der $x_{j0} = 1$ per definisjon og $t = s, s+1$. Ved forskjellige tidspunkter kan antallet varer være forskjellig. I det hele tatt kreves det ikke at varene må settes i par. En semi-log `dummy-variabel hedonisk funksjon` er gitt ved

$$\log p_j^t = \alpha^T x_j^t + \gamma_{t=s+1}\delta + \epsilon_j^t$$

der $\gamma_{t=s+1} = 1$ hvis $t = s+1$ og $\gamma_{t=s+1} = 0$ ellers, og $\alpha = (\alpha_0, \alpha_1, ..., \alpha_K)^T$ er kolonnevektoren med regresjonskoeffisienter. Betegn med $\hat{\delta}$ et estimat for δ. Den tilsvarende `dummy-variabel hedoniske indeks` er gitt ved

$$D^{s,s+1} = \exp(\hat{\delta}) \tag{37}$$

Man kan også bruke f.eks. den lineære dummy-variabel hedoniske funksjon gitt ved

$$p_j^t = \alpha^T x_j^t + \gamma_{t=s+1}\delta + \epsilon_j^t$$

der $\gamma_{t=s+1}$ og α defineres på samme måte som ovenfor. Den tilsvarende dummy-variabel hedoniske indeks er nå gitt ved

$$D^{s,s+1} = \hat{\delta}$$

4.3.2 Statistisk teori

Forhold til karakteristikk hedonisk funksjon En dummy-variabel indeks er uveid, og legger mindre krav på datagrunnlaget enn en karakteristikk indeks. Imidlertid kan en dummy-variabel hedonisk funksjon utledes fra tilsvarende karakteristikk hedoniske funskjoner, for tidspunkter s og $s+1$, ved at det innføres følgende restriksjoner

$$\alpha_0 = \beta_0^s \qquad \alpha_1 = \beta_1^s = \beta_1^{s+1} \qquad \cdots \qquad \alpha_K = \beta_K^s = \beta_K^{s+1} \qquad \alpha_0 + \delta = \beta_0^{s+1}$$

Slike restriksjoner betyr lite i praksis dersom $s+1$ og s ligger nær hverandre i tid, siden

$$\lim_{t \to s} \beta_k^t = \beta_k^s$$

Derfor anbefales metoden for *påfølgende* perioder (e.g. Triplett, 2004). På den andre side er det rett fram å definere en dummy-variabel hedonisk funksjon for flere enn to perioder. Hvorvidt restriksjonen på regresjonskoeffisienter er plausibel er i så fall et empirisk spørsmål.

Modell for endring En dummy-variabel hedonisk funksjon kan også motiveres som en modell for endring. Ta f.eks. semi-log spesifikasjon. For en vilkårlig vare j, har vi

$$\log p_j^{s+1} - \log p_j^s = \delta + \alpha^T (x_j^{s+1} - x_j^s) + \epsilon_j^{s,s+1} \qquad \text{der} \quad \epsilon_j^{s,s+1} = \epsilon_j^{s+1} - \epsilon_j^s$$

I tilfellet ingen endring i karakteristikk er den forventete prisendring gitt ved $E(p_j^{s+1}/p_j^s) \approx \exp(\delta)$, som gjelder for alle varer uten endring i karakteristikk. I tilfellet endring, har vi

$$E(p_j^{s+1}/p_j^s) \approx \exp(\delta) \exp\{\alpha^T (x_j^{s+1} - x_j^s)\}$$

der ekstra leddet kan tolkes som en justering som skyldes karakteristikkene som har endret seg. Koeffisientene α blir nå effekten av forskjellige karakteristikker på prisendringen. I tilfellet den hedoniske funksjon er definert for flere enn to tidspunkter, medfølges det en restriksjon om at denne effekt er konstant gjennom hele perioden.

Med identiske varer Anta først identiske varer $1, ..., n$ ved begge tidspunkter s og $s+1$, dvs. $x_j^t = x_j$. Den semi-log dummy-variabel hedoniske funksjon medfører at

$$\log I_j^{s,s+1} = \log p_j^{s+1} - \log p_j^s = \delta + (\epsilon_j^{s+1} - \epsilon_j^s)$$

Den eneste forskjell fra modell (10) bak Jevons indeksen er residualen. Under modell (10) betraktes p_j^s som konstant, mens her er den en stokastisk variabel som p_j^{s+1}. Vi kan sette δ lik μ_i i (10), eller ganske enkelt μ hvis vi dropper gruppeindikatoren. På denne måte har vi

$$E(\log P^{s,s+1}) = \delta \qquad \Rightarrow \qquad E(P^{s,s+1}) \approx \exp(\delta) = \exp(\mu) \approx E(D^{s,s+1})$$

Mao. Jevons indeksen og den semi-log dummy-variabel hedoniske indeks i dette tilfelle sikter på den samme teoretiske indeks, eller parameteren.

Med vareutskifting Anta nå at n byttes ut med n' ved tidspunkt $s+1$. En Jevons indeks basert på direkte imputering, dvs. $p_{n'}^{*s} = p_n^s$, er skjevt under modell (10), siden

$$E(\log P^{*s,s+1}) = \frac{1}{n} E\{\sum_{j=1}^{n-1} \log p_j^{s+1} + \log p_{n'}^{s+1} - \sum_{j=1}^{n-1} \log p_j^s - \log p_n^s\} = \mu + \frac{1}{n}(\log p_{n'}^s - \log p_n^s)$$

Under dummy-variabel hedoniske funksjonen har vi

$$E(\log P^{*s,s+1}) = \frac{1}{n} E\{\sum_{j=1}^{n-1} (\delta + \epsilon_j^{s+1} - \epsilon_j^s) + \delta + \alpha^T (x_{n'} - x_n) + (\epsilon_{n'}^{s+1} - \epsilon_n^s)\} = \delta + \frac{1}{n} \alpha^T (x_{n'} - x_n)$$

$$\Rightarrow \qquad E(D^{s,s+1}) \approx \exp(\delta) \approx E(P^{*s,s+1} / \exp\{\frac{1}{n} \alpha^T (x_{n'} - x_n)\}) \tag{38}$$

Mao. dummy-variabel hedoniske indeksen har omtrent den samme forventning som den justerte Jevons indeks basert på direkte imputering. Justeringen skyldes at vare n skiftes ut med n', med tilsvarende endringer i karakteristikkene.

Variansestimering En dummy-variabel hedonisk indeks er direkte en funksjon av den estimerte regresjonskoeffisient $\hat{\delta}$, betegnet med $D(\hat{\delta})$. F.eks. i tilfellet semi-log funksjon har vi $D(\hat{\delta}) = \exp(\hat{\delta})$, mens i tilfellet lineær funksjon har vi $D(\hat{\delta}) = \hat{\delta}$. En tilnærmet varians er gitt ved

$$V(D^{s,s+1}) = V\{D(\hat{\delta})\} \approx (\partial D/\partial \hat{\delta})^2 V(\hat{\delta})$$

Betegn med X designmatrisen under den aktuelle dummy-variabel hedoniske funksjon, hvis rader svarer til observasjonene og kolonner svarer til karakteristikkene. La y være tilsvarende kolonnevektoren av avhengige variabler, som er en funksjon av alle observerte priser. Sett $\xi^T = (\alpha^T, \delta)$, som har $K + 2$ komponenter. Vi har da

$$\hat{\xi} = (X^T X)^{-1}(X^T y) \qquad \Rightarrow \qquad Cov(\hat{\xi}, \hat{\xi}) = (X^T X)^{-1} X^T Cov(y, y) X (X^T X)^{-1}$$

En tilnærmet forventningrett sandwich variansestmator for $V(\hat{\delta})$ er da gitt ved

$$v = \phi^T \text{Diag}(\frac{e_i^2}{1 - h_i})\phi \qquad \text{der} \quad e = y - X\xi \quad \text{og} \quad H = X(X^T X)^{-1} X^T$$

og h_i er det $i$te element på diagonalen til matrise H, og ϕ^T er den siste rad i matrise $(X^T X)^{-1} X^T$.

4.4 Indekser basert på imputering

Betrakt situasjonen der vare n skiftes ut med n' fra tidspunkt s til $s+1$. Med en imputeringbasert hedonisk metode finner man først $p_{n'}^{*s}$ vha. en valgt hedonisk funskjon. Deretter kan en vanlig elementær indeksformel brukes på $(p_1^s, ..., p_{n-1}^s, p_{n'}^{*s})$ og $(p_1^{s+1}, ..., p_{n-1}^{s+1}, p_{n'}^{s+1})$.

4.4.1 Metoder

Dummy-variabel hedonisk imputering for endring Den justerte Jevons indeks i (38) kan omskrives som

$$P^{*s,s+1} = (\frac{p_{n'}^{s+1} \prod_{j=1}^{n-1} p_j^{s+1}}{p_n^s \prod_{j=1}^{n-1} p_j^s})^{\frac{1}{n}} / \hat{A}^{\frac{1}{n}} = \{\frac{p_{n'}^{s+1} \prod_{j=1}^{n-1} p_j^{s+1}}{(p_n^s \hat{A}) \prod_{j=1}^{n-1} p_j^s}\}^{\frac{1}{n}} \qquad \text{der} \quad \hat{A} = \exp\{\hat{\alpha}^T (x_{n'} - x_n)\}$$

Dette er en Jevons indeks med imputert $p_{n'}^{*s}$ gitt ved

$$p_{n'}^{*s} = p_n^s \hat{A} = p_n^s \exp\{\hat{\alpha}^T (x_{n'} - x_n)\}$$

Legg merke til at det er forskjellen mellom p_n^s og $p_{n'}^s$ som imputeres her. Fremgangsmåten er den samme for andre dummy-variabel hedoniske funksjoner, unntatt formen til justeringen. F.eks. i tilfellet lineær funksjon har vi

$$E(p_{n'}^s - p_n^s) = \alpha^T (x_{n'} - x_n) \qquad \Rightarrow \qquad p_{n'}^{*s} = p_n^s + \hat{\alpha}^T (x_{n'} - x_n)$$

Karakteristikk hedonisk imputering for endring Man kan også imputere forskjellen mellom p_n^s og $p_{n'}^s$ vha. en karakteristikk hedonisk funksjon, betegnet med

$$p_j^s = h(\beta^s, x_j^s, \epsilon_j^s)$$

der β^s og x_j^s er hhv. kolonnevektoren av regresjonskoeffisienter og karakteristikk, mens $h()$ varierer etter valget av funksjon. F.eks. i tilfellet semi-log karakteristikk hedonisk funksjon, har vi

$$p_{n'}^s = h(\beta^s, x_{n'}, \epsilon_{n'}^s) = \exp(x_{n'}^T \beta^s + \epsilon_{n'}^s)$$

Det følger at

$$E(\log p^s_{n'} - \log p^s_n) = (x_{n'} - x_n)^T \beta^s \qquad \Rightarrow \qquad p^{*s}_{n'} = p^s_n \exp\{(x_{n'} - x_n)^T \hat{\beta}^s\}$$

der $\hat{\beta}^s$ inneholder estimerte regresjonskoeffisientene. Den eneste forskjell fra dummy-variabel hedoniske imputeringen er regresjonskoeffisientene, som nå er $\hat{\beta}^s$ istedet for $\hat{\alpha}$. I praksis forventer man liten forskjell mellom de to dersom s og $s+1$ ligger nær hverandre i tid.

Direkte karakteristikk hedonisk imputering Istedenfor å imputere forskjellen mellom $p^s_{n'}$ og p^s_n, kan man imputere $p^{*s}_{n'}$ direkte basert på karakteristikk hedoniske funksjonen, nemlig

$$p^{*s}_{n'} = h(\hat{\beta}^s, x_{n'}, 0)$$

der residualen $\epsilon^s_{n'}$ settes til 0, dvs. sin forventning. I den senere tid har det også blitt forslått en *dobbel* karakteristikk hedonisk imputeringsmetode, dvs.

$$p^{*s}_{n'} = h(\hat{\beta}^s, x_{n'}, 0) \qquad \text{og} \qquad p^{*s+1}_{n'} = h(\hat{\beta}^{s+1}, x_{n'}, 0)$$

og den imputerte $P^{*s,s+1}$ beregnes basert på $(p^s_1, ..., p^s_{n-1}, p^{*s}_{n'})$ og $(p^s_1, ..., p^s_{n-1}, p^{*s+1}_{n'})$.

Direkte imputering eller imputering for endring? Det kan virke rart å kaste bort den observerte $p^{s+1}_{n'}$ til fordel for den imputerte $p^{*s+1}_{n'}$ med dobbel imputering. Triplett (2004) diskuterte begrunnelser som har blitt forslått. Vår forklaring er som følger. Legg først merke til at direkte karateristikk imputering $p^{*s}_{n'} = h(\hat{\beta}^s, x_{n'}, 0)$ ikke inneholder en residual. Det oppstår dermed en viss ubalans i forholdet til $p^{s+1}_{n'}$, som inneholder residual $\epsilon^{s+1}_{n'}$. Dobbel imputeringen gjenoppretter balansen ved å bruke $p^{*s+1}_{n'}$, som heller ikke inneholder residual. Men hvorfor bytter man ikke ut alle observerte prisene med imputeringer? Vi mener at svaret ligger i robusthet overfor misspesifikasjon av modellen: man ville ikke har brukt den hedoniske metode og byttet ut alle observasjonene med imputeringer, dersom man også hadde observert $p^s_{n'}$. Legg videre merke til at, i tilfellet imputering for endring, $p^{*s}_{n'}$ avhenger av ϵ^s_n via p^s_n. Det er generelt uklart hva som er best, dobbel imputering uten residualer eller imputering for endring med $\epsilon^{s+1}_{n'}$ og $\epsilon^{*s}_{n'} = \epsilon^s_n$, så lenge det gis rom for misspesifikasjon av modellen.

Utskifting av karakteristikk Hva skjer hvis vare n' inneholder karakteristikk som ikke finnes ved tidspunktet s? En dummy-variabel hedonisk funksjon er ikke lenger plausibel, da det er urimelig å anta at innføringen av en ny karakteristikk ikke påvirker 'prisene' på eksisterende karakteristikk. Det er umulig å imputere $p^{*s}_{n'}$ vha. en karakteristikk hedonisk funksjon siden prisen for den nye karakteristikk ikke kan estimeres ved tidspunkt s. Det er mulig å imputere p^{*s+1}_n og bruke (p^s_n, p^{*s+1}_n) istedenfor $(p^{*s}_{n'}, p^{s+1}_{n'})$, under forutsetningen at vare n ikke har karakteristikk som er utgått ved tidspunkt $s+1$. Plausibelt er det likevel ikke siden den nye karakteristikk ikke kommer med i indeksberegningen, til tross at den er med på å bestemme karakteristikk prisene ved tidspunktet $s+1$. En direkte karakteristikk hedonisk indeks er i teori et bedre alternativ.

4.4.2 Bootstrap variansestimering

Variansestimering for hedoniske indekser basert på imputering er noe mer komplisert pga. at de imputerte priser må behandles annerledes enn de som er direkte observert. Generisk betegn med $p^{*t}_j = g(\hat{\xi})$ imputeringen for vare j ved tidspunktet t. Funksjonen $g()$ avhenger av den aktuelle hedoniske funksjon, der ξ inneholder relevante regresjonskoeffisientene og $\hat{\xi}$ betegner estimatene deres. I tillegg avhenger $g()$ av karakteristikk. For enkelhetsskyld tas ikke de med i notasjonen.

For å gjøre variansen mer sammenlignbar med variansen for en fastvektsindeks, skal vi betrakte basisprisene som konstante. Kun direkte observerte priser ved statistikkperioden er stokastiske, betegnet med p_j^t for $j = 1, ..., n$, dvs. uten å gjøre forskjell på eventuelle nye og gamle varer. Under en hedonisk funksjon, kan variasjonen i data oppsummeres i estimerte residualer gitt ved

$$e_j^t = h^{-1}(p_j^t) - h^{-1}(\hat{\xi}, x_j^t, 0)$$

der $h^{-1}()$ betegner invers funksjon til $h()$. F.eks. i tilfellet semi-log karakteristikk funksjon, dvs. $\log p_j^t = x_j^T \beta^t + \epsilon_j^t$, har vi

$$h^{-1}(p_j^t) = \log p_j^t \qquad \text{og} \qquad h^{-1}(\hat{\xi}, x_j^t, 0) = h^{-1}(\hat{\beta}^t, x_j^t, 0) = x_j^T \hat{\beta}^t$$

Vi skal nå beskrive en bootstrap metode for variansesitmering. Den er basert på en standard ikke-parametrisk bootstrap metode for regresjonsmodellene (e.g. Davison and Hinkley, 1997).

- Trekk tilfeldig og med tilbakelegging $\epsilon_j^{'t}$ blant $\{e_1^t, ..., e_n^t\}$, for $j = 1, ..., n$.
- Sett $p_j^{'t} = h(\hat{\xi}, x_j, \epsilon_j^{'t})$, for $j = 1, ..., n$, som avhenger av den akutelle hedoniske funksjon.
- Estimer $\hat{\xi}'$ basert på alle oberverte p_j^s og $(p_1^{'t}, ..., p_n^{'t})$ som vanlig.
- Imputer $p_j^{*s} = g(\hat{\xi}')$ for alle som mangler priser ved tidspunkt s.
- Beregn $P^{'s,t}$ basert på alle observerte p_j^s og imputerte p_j^{*s} og $(p_1^{'t}, ..., p_n^{'t})$ som vanlig.

Vi har nå fått $P^{'s,t}$ som et *bootstrap replikat* for den opprinnelige imputerte $P^{*s,t}$. Uavhengige gjentagelser av opplegget gir oss et sett av uavhengige bootstrap replikater, betegnet med $P_{(1)}^{'s,t}, ..., P_{(B)}^{'s,t}$ der B er antallet gjentagelser. Et bootstrap variansestimat for $V(P^{*s,t})$ er da gitt ved

$$v_{boot}(P^{*s,t}) = \frac{1}{B-1} \sum_{b=1}^{B} \{P_{(b)}^{'s,t} - \frac{1}{B} \sum_{l=1}^{B} P_{(l)}^{'s,t}\}^2$$

4.5 Eksempel: Boligprisindeks

To problemer To problmer med Boligprisindeksen (BPI) gjør at mange tar i bruk hedoniske metoder. Det første problem er knyttet til det faktum at det aldri selges samme bolig to ganger. F.eks. de aller fleste boliger som omsettes i to påfølgende kvartaler er forskjellige. Heller ikke er en bolig den samme når den selges på nytt: den er annerledes uansett om den er blitt oppusset i mellomtiden eller ikke. Mangelen på identiske varer over tid, og dermed grunnlaget for en fastvektsindeks, er tydelig i BPI. Det andre problem dreier seg om forholdet mellom utvalg og populasjon. Kun en omsatt bolig har en pris. Men salgsprosessen kan ikke kontrolleres av en utvalgsplan, enten sannsynlighetsbasert eller ikke. På hvilken måte er en observert solgt bolig representativ, dersom man er interessert i en prisindeks for hele boligmassen? I tillegg er begrepet om en prisindeks for hele boligmassen noe uklart. Tenk om hva ville ha skjedd med prisen dersom alle boliger settes ut til slags på en gang?

Karakteristikk metode En karakteristikk hedonisk funksjon for tidspunktet t kan f.eks. være

$$\log p_j^t = (x_j^t)^T \beta^t + \epsilon_j^t \qquad \text{der} \quad x_{j0} = 1$$

Man kan bruke andre avhengige variabler som f.eks. pris per kvm. Vanlige uavhengige variabler er areal (på log-skala), beliggenhet (f.eks. region eller postnummer), antall bad og/eller rom, osv. Legg merke til at en variabel som beliggenhet må spesifiseres som en 0/1-vektor. Anta det finnes q kategorier for beliggenhet. Da består den tilsvarende 0/1-vektor av q komponenter, som alle er lik 0 unntatt den som svarer til den aktuelle beliggenhetskategori, og den settes til 1.

En observert bolig blir nå representativ under den hedoniske antagelse, nemlig at de observerte karakteristikker er representative for hele boligmassen. På samme måte er det uproblematisk at forskjellige boliger observeres ved forksjellige tidspunkter. Imidlertid inneholder den hedoniske funksjon misspefisikasjoner som generelt gjelder. Ta f.eks. beliggenheten som kanskje er den meste sentrale prisfaktor for en bolig. Man kan med rett hevde at beliggenheten er unik for hver bolig, dersom det tas i betraktning alle fysiske og miljømessige sider som kan påvirke prisen. Det er også lett å komme på prisfaktorer som ikke er tatt med i enhver gitt funksjon pga. mangel på data, eller at karakteristikken ikke er fysisk målbar i det hele tatt.

Mangel på data medfører også vanskeligheter i valget av indeksformel. Populasjonsvektene er umulige i praksis, så lenge man ikke har god nok registering av boligmassen med relevante karakteristikker. Alternativet er å beregne indeksen for en standard (eller representativ) bolig. Men hvordan skal man f.eks. karakterisere en standard beliggenhet? Å dele boligmassen inn i strata for beliggenhet er vel nær sagt det eneste man kan gjøre. Imidlertid kan man ikke definere for mange av dem heller, da det lett kan bli for få observasjoner i hvert stratum.

Dummy-variabel metode En semi-log dummy-variabel hedonisk funksjon kan f.eks. være

$$\log p_j^t = \alpha^T x_j^t + \sum_{u=s}^{s'} \delta_u \gamma_{t=u} + \epsilon_j^t \qquad \text{for} \quad t = s, s+1, ..., s'$$

Det er vanlig å bruke flere enn to påfølgende perioder for å akkumulere prisobservasjoner. Indeksen $D^{s,t}$ er kun en funksjon av $\hat{\delta}_t - \hat{\delta}_s$, selv om også alle andre δ_u estimeres. I tilfellet antallet perioder (dvs. $s' - s + 1$) holdes fast, forekommer kjeding ved hvert tidspunkt, da prisbasisperioden endres hele tiden. Dummy-variabel metoden medfører at indeksen beregnes for en standard bolig, da

$$\frac{E(p_x^t)}{E(p_x^s)} \approx \frac{\exp(\hat{\alpha}^T x + \delta_t)}{\exp(\hat{\alpha}^T x + \delta_s)} = \exp(\hat{\delta}_t - \hat{\delta}_s)$$

uten at man trenger å spesifisere karakteristikkene (betegnet med x ovenfor) til standard boligen. Det er mange som bruker dummy-variabel metoden for å beregne BPI.

Gjentatt-salg metode For en identisk bolig medfører en semi-log dummy-variabel funksjon at

$$\log p_j^t - \log p_j^s = (\delta_t - \delta_s) + (\epsilon_j^t - \epsilon_j^s)$$

dvs. man trenger ikke en gang å spesifisere karakteristikkene. Det er ikke bare lettvint i praksis, men har teoretiske fordeler siden en karakteristikk funksjon generelt er misspesifisert. Det mest nærliggende valg på identisk bolig er selvsagt en bolig som selges igjen. Det er mange som bruker denne såkalte gjentatt-salg metode.

En kritikk av metoden går ut på problemet mellom utvalg og populasjon, da man føler at boliger som selges igjen og igjen må være litt annerledes enn boliger som ikke selges like ofte. Det man vet imidlertid er f.eks. at eneboliger selges sjeldnere enn leiligheter, og boliger på landet selges sjeldnere enn boliger i byen, ol. Vanskeligheten er til en viss grad overkommelig ved å innføre strata (eller segmenter) for bolig, som nettopp er rettet mot hyppigheten i gjentatt-salg. Et praktisk problem er at det kan bli få prisobservasjoner i et stratum med lav salgshyppighet. Husk at dette kommer på toppen av at gjentatt-salg i utgangspunkt er en undermengde av alle boligsalg.

En annen kritikk går ut på at det aldri selges samme bolig to ganger, som nevnt før. Oppussing i mellomtiden må tas med i betraktning, eller slitasje i det motsatte fall. Goetzmann and Spiegel (1995) forslo en justering ved å innføre en konstant i modellen, dvs.

$$\log p_j^t - \log p_j^s = \lambda + (\delta_t - \delta_s) + (\epsilon_j^t - \epsilon_j^s)$$

der λ står for en gjennomsnittlig prisendring mellom gjentatt-salg, utenom den *rene* prisendring over tid. Et spørsmål er da om λ skal tillegges boliger som ikke selges så ofte. Legg merke til at modellen ikke er identifiserbar med kun to tidspunkter da pga. kolinearitet mellom λ og $\delta_{s+1} - \delta_s$.

Generalisert fastvektsindeks Det er fremdeles mange som bruker denne metode uten at den er kjent under en slik betegensele. Kort sagt innnfører man et sett av strata for boliger med homogen prisutvikling, betegnet med $h = 1, ..., H$. Disse fungerer som elementære grupper for en fastvektsindeks. Betegn med n_h^t antallet boliger som observeres ved tidspunkt t, og lignende for n_h^s. Da kan man beregne en *generalisert* Jevons indeks for hvert stratum ved

$$P_h^{s,t} = \frac{(\prod_{j=1}^{n_h^t} p_{hj}^t)^{1/n_h^t}}{(\prod_{j=1}^{n_h^s} p_{hj}^s)^{1/n_h^s}} \tag{39}$$

Den er en generalisering siden $n_h^s \neq n_h^t$. Varene trenger heller ikke å være identiske ved begge tidspunkter. Den underliggende statistiske modell er nå gitt ved

$$\log p_{hj}^t = \mu_h^t + \epsilon_{hj}^t \qquad \text{der} \quad E(\epsilon_{hj}^t) = 0 \quad \text{og} \quad V(\epsilon_{hj}^t) = \psi_h^t$$

dvs. med tidsbestemt μ^t som det generelle prisnivå, som estimeres med observerte geometriske gjennomsnittet av priser. Den teoretiske elementære indeks er $\theta_h^{s,t} = \exp(\mu_h^t - \mu_h^s)$.

Generalisert fastvektsindeks med hedonisk justering for karakteristikk Selv når man godtar modellen under f.eks. den generaliserte Jevons indeks (39), kan det oppstår problemer pga. til dels store forskjeller i utvalgssammensetningen over tid. I grunn skyldes det mangelen på priser for samme bolig over tid. Anta for enkelhetsskyld at det finnes kun ett stratum, og betrakt den generaliserte Jevons indeks $P^{s,t}$ som da er gitt ved (39). En justering for forskjellen i utvalgssammensetning kan fås vha. dummy-variabel hedonisk imputering, dvs.

$$P^{*s,t} = P^{s,t} / \exp\{\hat{\alpha}^T(\bar{x}^t - \bar{x}^s)\} = \frac{\exp\{\frac{1}{n^t}\sum_{j=1}^{n^t} \log p_j^t - \hat{\alpha}^T \bar{x}^t\}}{\exp\{\frac{1}{n^s}\sum_{j=1}^{n^s} \log p_j^s - \hat{\alpha}^T \bar{x}^s\}} \tag{40}$$

der $\bar{x}^t = \frac{1}{n^t}\sum_{j=1}^{n^t} x_j^t$ og $\bar{x}^s = \frac{1}{n^s}\sum_{j=1}^{n^s} x_j^s$ hhv. er gjennomsnittlig karakteristikk i utvalget.

Referanser

Allen, R.G.D. (1963). Price index numbers (with discussion). *International Statistical Review*, **31**, 281–306.

Balk, B.M. (2005). Price indexes for elementary aggregates: The sampling approach. *Journal of Official Statistics*, **21**, 675–699.

Clements, K.W., Izan, H.Y., and Selvanathan, E.A. (2006). Stochastic index numbers: A review. *International Statistical Review*, **74**, 235–270.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press.

Edgeworth, F. Y. (1925). *Several articles collected in* Papers Relating to Political Economy*, Volume I.* London.

Fisher, I. (1922). *The Making of Index Numbers.* Boston.

Frisch, R. (1936). Annual survey of general economic theory: The problem of index numbers. *Econometrica*, **4**, 1–38.

Goetzmann, W.N. and Spiegel, M. (1995). Non-temporal components of residential real estate appreciation. *Review of Economics and Statistics*, **14**, 199–206.

International Labour Office (2004). *Consumer Price Indices: An ILO Manual.* Geneva.

Krönus, A.A. (1924). The problem of the true index of the cost of living. *English translation, 1939, Econometrica*, **7**, 10–29.

Longva, S. (1971). *Økonomiske Indekstall: Prinsipper og beregningsmetoder.* Universitetsforlaget.

Selvanathan, E.A. and Prasada Rao, D.S. (1994). *Index Numbers: A Stochastic Approach.* London: Macmullan.

Solheim, L. (2004). Indeksberegninger: Prinsipper, teori, metoder og eksempler. Upublisert notat, Statistisk Sentralbyraa.

Theil, H. (1867). *Economics and Information Theory.* Amsterdam and Chicago: North-Holland and Rand McNally.

Triplett, J. (2004). Handbook on hedonic indexes and quality adjustments in price indexes: Special Application to information technology products. OECD, STI Working Paper 2004/9.

Valliant, R. (1992). Smoothing variance estimates for price indexes over time. *Journal of Official Statistics*, **8**, 433–444.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). *Finite Population Sampling and Inference.* New York: Wiley.

De sist utgitte publikasjonene i serien Notater

2006/43 A-G. Jørstad: Overvåkingssystemet for bedrifter i Bof. 19s.

2006/44 M. Høstmark og B.O. Lagerstrøm: Undersøkelse om Arbeidsmiljø: Destruktiv atferd i arbeidslivet. Dokumentasjonsrapport. 43s.

2006/45 T.K. Schjerven og K.Å. Wass: Faglig modell og rammeverk i StatRes. 67s.

2006/46 R. Sønsterudbråten: FOB2001. Dokumentasjon av logistikk og svartjeneste. 68s.

2006/47 K. Henriksen: Utvalgsplan til konsumprisindeksens nye matvareindeks - Basert på strekkodedata. 23s.

2006/48 A.B. Thorud, D. Rafat, S. Ferstad og E. Vinju: Tverrgående revisjon i KOSTRA - Bedring av påliteligheten i nøkkeltallene. 65s.

2006/49 T. Granseth: Grensehandel. En analyse av kvaliteten av data. 48s.

2006/50 E. Engelien, H. Høie og M. Steinnes: Bygging i strandsona. Metode og resultater. 18s.

2006/51 A. Akselsen, K.I. Bøe og Ø. Sivertstøl: FD - Trygd. Dokumentasjonsrapport. Arbeidssøkere, 1.1.1992-30.4.2001. 75s.

2006/52 L. Østby: Bruk av velferdsordninger blant nyankomne innvandrere fra de nye EØS-landene i 2005. 34s.

2006/53 G. Claus: Inntekts- og formuesundersøkelsen for personlig næringsdrivende 2004. Dokumentasjon. 28s.

2006/54 J. Heldal: Logistisk regresjon - kurskompendium i byråskolens kurs SM507. 51s.

2006/55 L.H. Thingstad: Varehandelsstatistikk 2002 - omsetning etter varegruppe. 59s.

2006/56 H.Kull Brofoss og A. Barstad: Internasjonale erfaringer med områderettede tiltak i storbyer. En litteraturstudie. 101s.

2006/57 B. Bye og I. Ringdal: Disaggregering av helse-, omsorg- og utdanningstjenester i MSG6-modellen. 39s.

2006/59 Leiemarkedsundersøkelsen 2006. Dokumentasjonsrapport. 43s.

2006/60 J. Hamre og A. Vedø: Utvalgsundersøkelse om egenmeldt sykefravær. Dokumentasjon av utvalgsplanen, utvalget for 2006 og standardfeilberegninger. 50 s.

2006/61 E. C. Rauan: Undersøking om foreldrebetaling i barnehagar, august 2006. 45s.

2006/62 Indikatorer på kjemikalieområdet - Risiko for skade på helse og miljø grunnet bruk av kjemiske stoffer, fase 2. 100s.

2006/63. Lønnsstatistikk 2006. Etablering av populasjon og utvalg. Dokumentasjonsnotat. 51s.

2006/64. Bygg, anlegg og eiendomsdrift - tall og metode. 53s.

2006/65: O. Villund: Forsøk med imputering av utførte timeverk i Arbeidskraftundersøkelsen. 58 s.

2006/66. FD - Trygd Dokumentasjonsrapport. Arbeidssøkere 1.5.2001-31.12.2004. 60s.

2006/67: E. Holmøy: Non-Ponzi-Game betingelser og lukking av anvendte intertemporale likevektsmodeller. 38s.

2006/68. Kirkelig rapportering 2006 Felles- og menighetsråd. 19s.

2006/69. FD-Trygd Dokumentasjonsrapport. Stønader til enslig forsørger. 1992-2005. 45s.

2006/70. Imputering i AKU for undersysselsetting. 19s.

2006/71. Gruppering av kommuner for kontroll av yrkesdata. 25s.