

Observationer av webbskrapade data

För information

Prisenheten har sedan sommaren 2019 genom webbskrapning veckovis samlat in data från en prisjämförelsesida. Insamlingen har avgränsats till vissa produktgrupper i Konsumentprisindex (KPI) och till information om pris, märke, modell, popularitet och lagerstatus. På detta underlag har vissa analyser kunnat göras som del i ett projekt med finansiering från Eurostat (s.k. Granted project). I denna promemoria redogörs kortfattat för dessa analyser, och indexberäkningar baserade på underlaget presenteras. Preliminärt rapportunderlag från analysen biläggs.

Sammanfattning av dataanalysen

Genom så kallad *webbskrapning* från en välkänd prisjämförelsesida har prisuppgifter för fyra typer av elektronikprodukter som ingår i KPI som produktgrupper samlats in veckovis sedan sommaren 2019: *mobiltelefoner, datorer, tv-apparater* och *tvättmaskiner*. Insamlingen har gjorts veckovis och analyserats. Inga betydande dataproblem har noterats från insamlingen. Följande aspekter har kommit upp i analysen.

- Tillgänglighet: lagerstatus anges och avgränsar därmed vad som faktiskt finns tillgängligt för köp, enligt visad information
- Butikstyp: endast nätbutik eller butik med fysisk närvaro.
- Variant: inom ett märke och modell finns en del variation på specifikationer.
- Livslängd på marknaden: hur länge finns produkterbjudandena och produkterna tillgängliga? En begränsning för studien är tidsspannet som webbskrapningen gjorts (upp till 11 månader i denna analys).
- Utbud: skiljer sig tydligt mellan butiker. Tillfälliga produkterbjudanden kan utgöra en betydande del av utbudet.
- Relativa priser: hur förhåller sig priserna till varandra över tid?
- Försäljningskvantiteter finns inte att tillgå. Slutsatser dras därför enbart på nominella priser/uppträdanden på hemsidor och inte vidare än så.

Dataanalysen kan sammanfattas enligt följande.

Många produkterbjudanden har tämligen kort livslängd (s.k. *duration*) och ibland saknas de temporärt eller är ej tillgängliga med hänsyn till lagerstatus, medan andra produkterbjudanden kan observeras över längre tid.

Märke är mer betydande än butik för att förklara pris och prisförändringar. Som förväntat är produkten som sådan en bra prediktor för dess prisnivå, oavsett butik.

Att välja den typ av butiker som har både fysisk och online försäljning bidrar inte tydligt mycket mer till att förklara pris eller prisförändringar. Däremot har den senare typen, välkända butiker med sådan dual närvaro, i allmänhet mindre prisvariation per produkt och betydligt större sortiment än de butiker som erbjuder enbart onlineförsäljning.

Det finns tydliga tendenser att lägre priser fortsätter att vara låga, i relativa termer, samtidigt som högre priser tenderar att fortsätta vara höga. Detta är ett aggregerat resultat men det indikerar att butiker/prissättning skiljer sig åt i något avseende vilket bör beaktas i urvalssammanhang.

Indexberäkningar

Som en del av dataanalysen av det webbskrapande materialet har även alternativa prisindex beräknats som en fallstudie för två utvalda produktgrupper som månatlig kedjning tillämpas för i KPI.

Analysens syfte

I analytiskt syfte har vi använt data från en prisjämförelsesida och beräknat prisindex och jämfört med det index som används i KPI. Genom denna analys undersöker vi om SCB:s urval skiljer sig från alternativa urval från webbskrapade data i termer av *prisförändring* liksom effekter av *kedjning*. Målsättningen är att se eventuella effekter av *urvalet* och därför prövas olika korgar.

Studiens genomförande

Alternativa prisindex har beräknats för datorer och mobiltelefoner baserat på det webbskrapade materialet. Genom att låta vissa parametrar i undersökningsdesignen variera, har olika indexutfall beräknats. Med fem olika parametrar i undersökningsdesignen har sammanlagt 690 olika prisindex beräknats. De parametrar som analyseras är indextyp, butiksurval, lägsta pris, aggregering över butiker och urval av popularitet.

Tabell 1: Beskrivning av olika parametrar vid undersökningsdesign

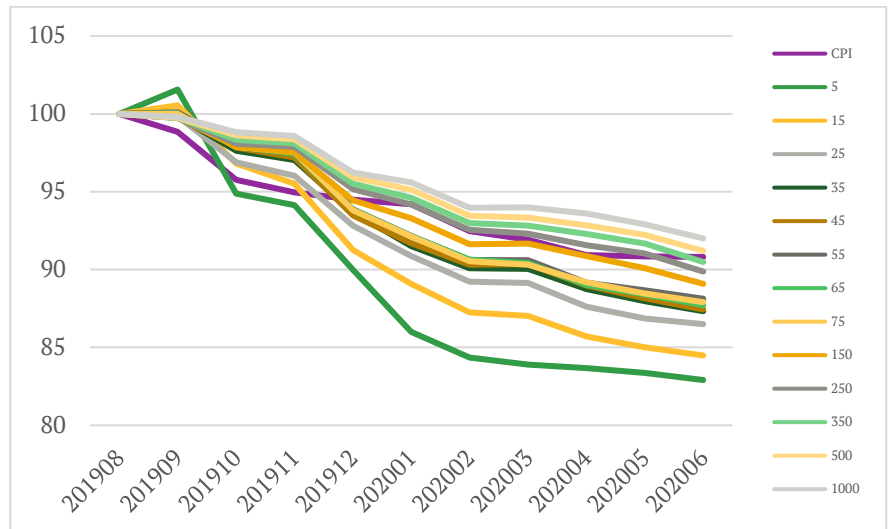
| Parameter | Parametervärden | Antal parameter-värden | Beskrivning |
|---------------------|---|------------------------|---|
| Indextyp | Månadskedjning Enhetsvärdeindex | 2 | Månadskedjning av oviktat Jevons index eller Enhetsvärdeindex |
| Butiksurval | Stort urval Litet urval | 2 | Två olika urval av butiker. <i>Litet urval</i> är en delmängd av det stora urvalet |
| Prismättnings metod | Max-priset Min-priset Medelpriset | 3 | Prisnotering av det lägsta priset, medelpriset eller högsta noterade priset inom en månad |
| Produkt-definition | Produkt Produkt*butik | 2 | Definition av prisnoteringens objekt. Antingen en prisnotering per produkt, eller en prisnotering per produkt i butik |
| Popularitet | 5,10,15,...,400, 500, 700,1000 | 29 | Urval av antalet objekt. De topp "n" (antal) produkter enligt indikerad popularitet . |

Resultat

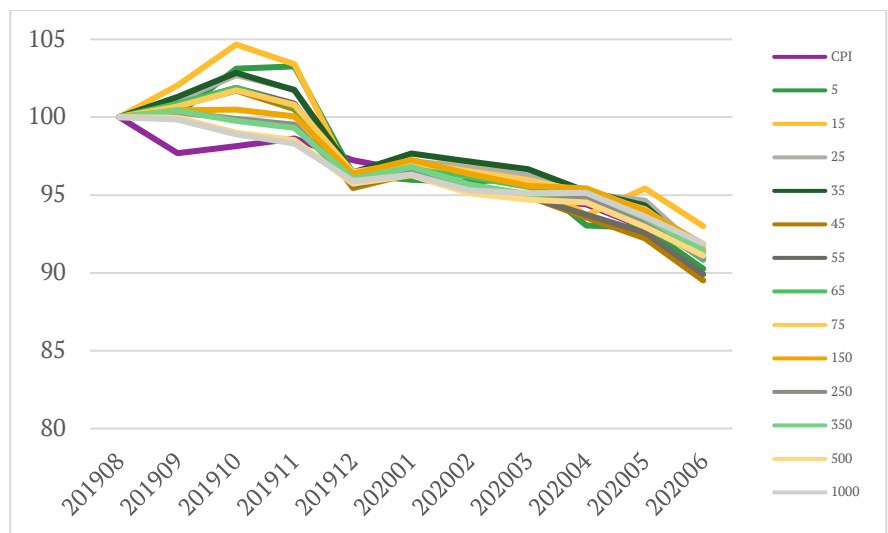
I följande avsnitt redovisas jämförelser av indexberäkningar för varje parameter för sig. Vid jämförelse av en parameter är övriga parametrar satta till att efterlikna den designen som råder i det officiella delindexet i KPI. För att kunna särskilja priset och kedjningseffekt på prisindex, har även enhetsvärdeindex beräknats utan kedjning av produkter, där den allmänna prisnivån för aktuell korg representerar indexutvecklingen. Ett enhetsvärdeindex skildrar enbart snittpriset för varorna i korgen och tar inte hänsyn till ändrad kvalitet i korgen, således är det inget index som vanligtvis används i beräkningen av KPI. Alla observationer i korgen är enhetligt viktade i beräkningarna, då det inte finns tillgång till försäljningskvantiteter i datamaterialet.

Popularitet

En cut-off har tillämpats implicit genom att använda popularitet. Eftersom det inte finns tillgång till försäljningskvantiteter, är ett alternativ till produkturval att prismäta de mest populära produkterna.

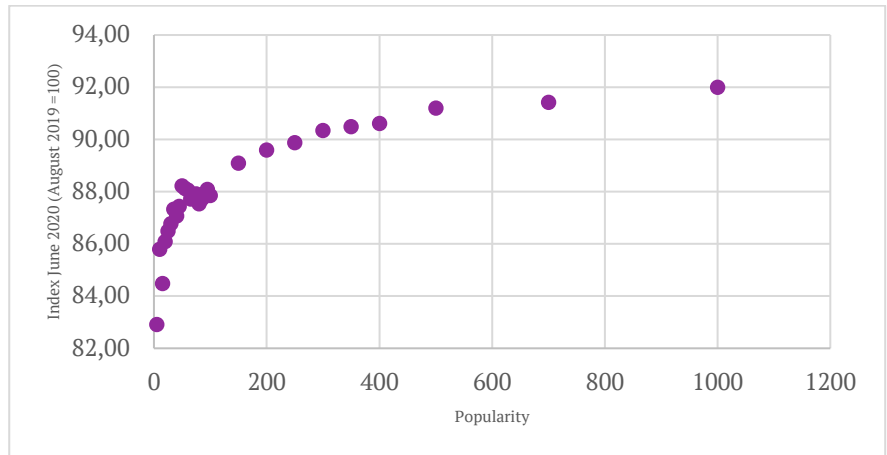


Figur 1: Månadskedjning för mobiltelefoner vid olika storlekar av produkturval baserat på popularitet (augusti 2019=100)

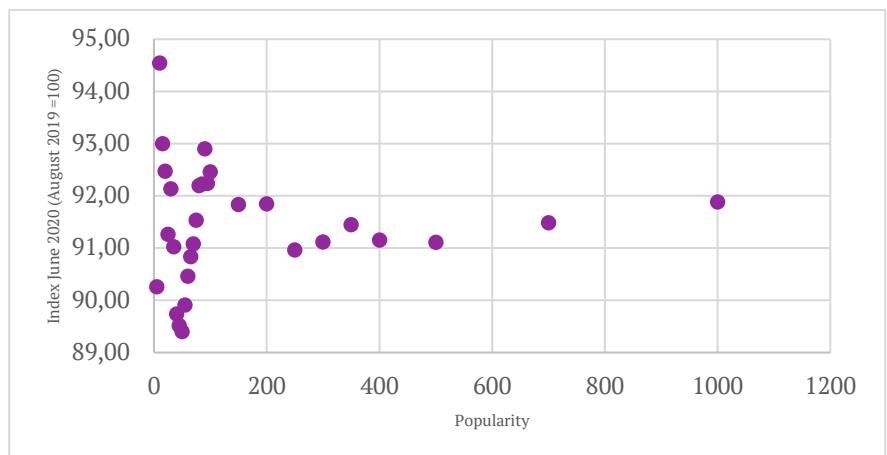


Figur 2: Månadskedjning för datorer vid olika storlekar av produkturval baserat på popularitet (augusti 2019=100)

Mobiltelefoner och datorer visar olika karaktäristika vad gäller storleken av antal mest populära produkter som index beräknas på. Mobiltelefoner har en större spridning på index, medan index för datorer följer varandra mer över tid.

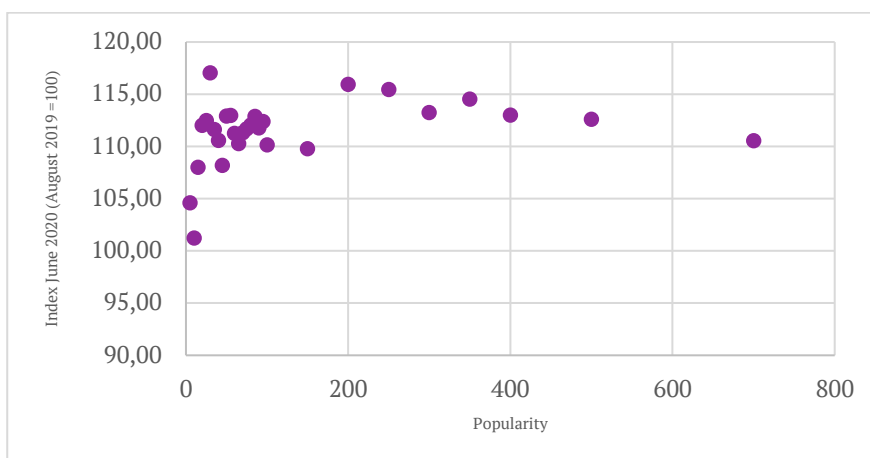


Figur 3: Spridning av indexutfall juni 2020 för månadskedjning av mobiltelefoner vid olika storlekar av produkturval baserat på popularitet (augusti 2019=100)

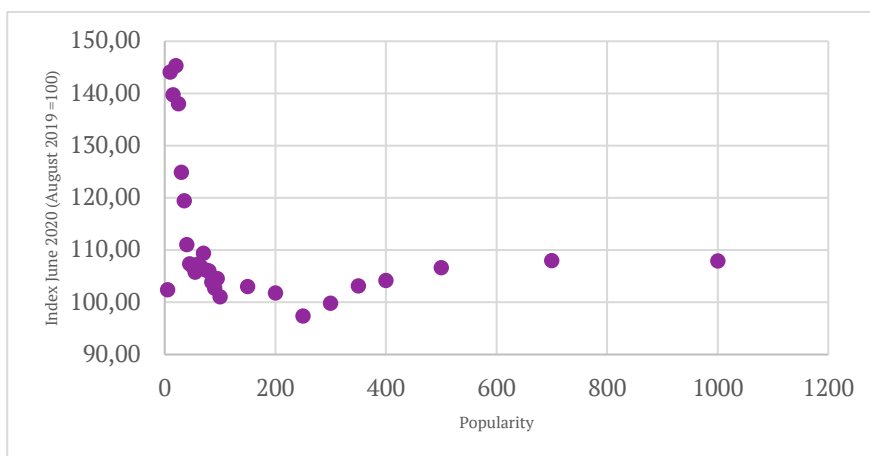


Figur 4: Spridning av indexutfall juni 2020 för månadskedjning av datorer vid olika storlekar av produkturval baserat på popularitet (augusti 2019=100)

I figur 3 och 4 visas spridningen av indexutfall av månadskedjning av mobiltelefoner och datorer baserat på storleken av antal mest populära produkter. För mobiltelefoner framgår det tydligt att det finns en positiv korrelation mellan indexutfall och storlek på produkturvalet. Vid månadskedjning antas implicita kvalitetsvärderingar varje månad, och indexutfall kan både bero på de implicita kvalitetsvärderingarna samt på ren prisseffekt.



Figur 5: Spridning av indexutfall juni 2020 för enhetsvärdeindex av mobiltelefoner vid olika storlekar av produkturval baserat på popularitet (augusti 2019=100)

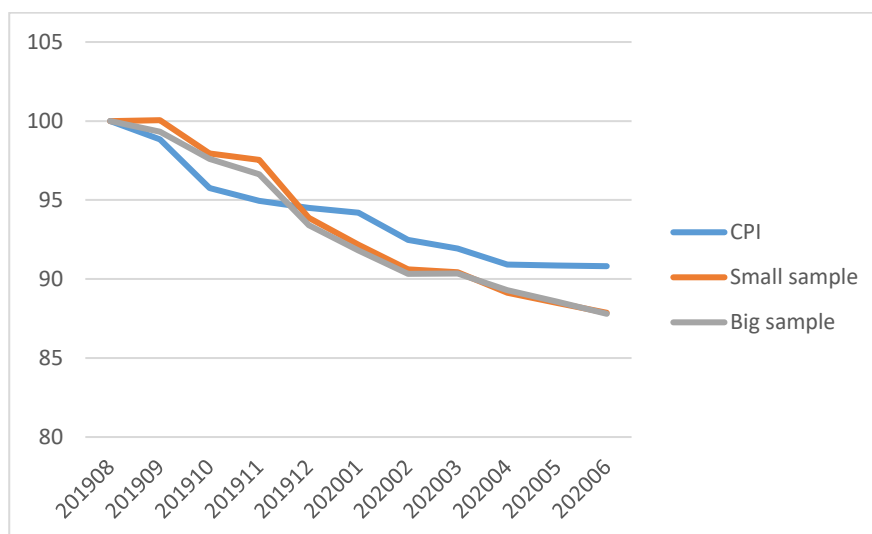


Figur 6: Spridning av indexutfall juni 2020 för enhetsvärdeindex av datorer vid olika storlekar av produkturval baserat på popularitet (augusti 2019=100)

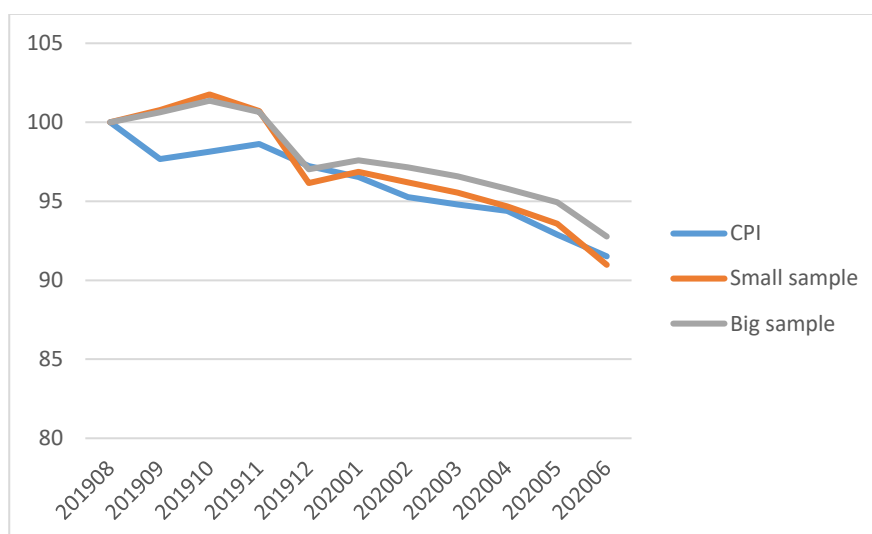
Figur 5 och 6 visar spridning av indexutfall vid beräkning av enhetsvärdeindex. Vid beräkning av enhetsvärdeindex jämförs enbart den genomsnittliga prisnivån av produkterna i månadens korg. För datorer verkar prisnivån ha utvecklats högre vid en liten korg av populära produkter. För mobiltelefoner verkar detta samband visuellt vara det omvända.

Butiksurval

Vid urval av antal butiker har två olika urval gjorts. Ett litet urval som innehåller de största återförsäljarna av mobiltelefoner respektive datorer, samt ett större urval där flera mindre företag också är inkluderade.



Figur 7: Månadskedjning av mobiltelefoner vid olika storlekar av butiksurval (augusti 2019=100)

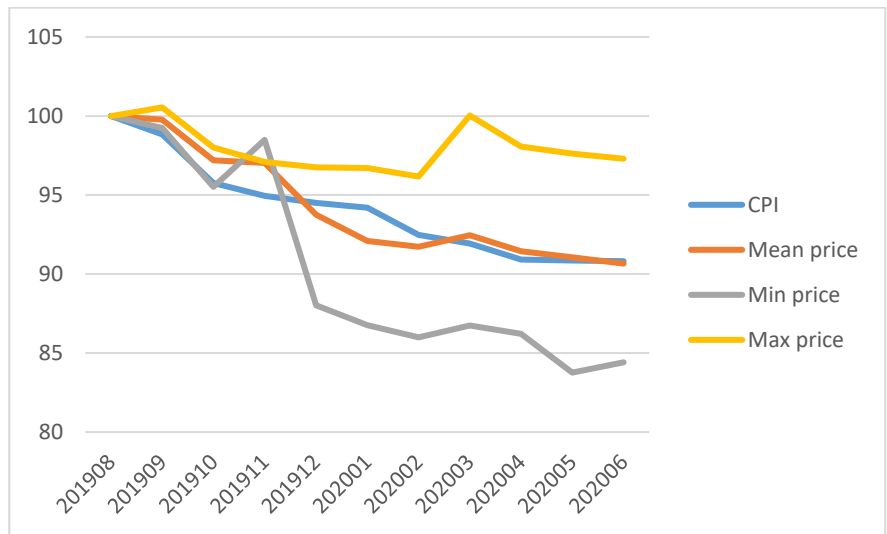


Figur 8: Månadskedjning av datorer vid olika storlekar av butiksurval (augusti 2019=100)

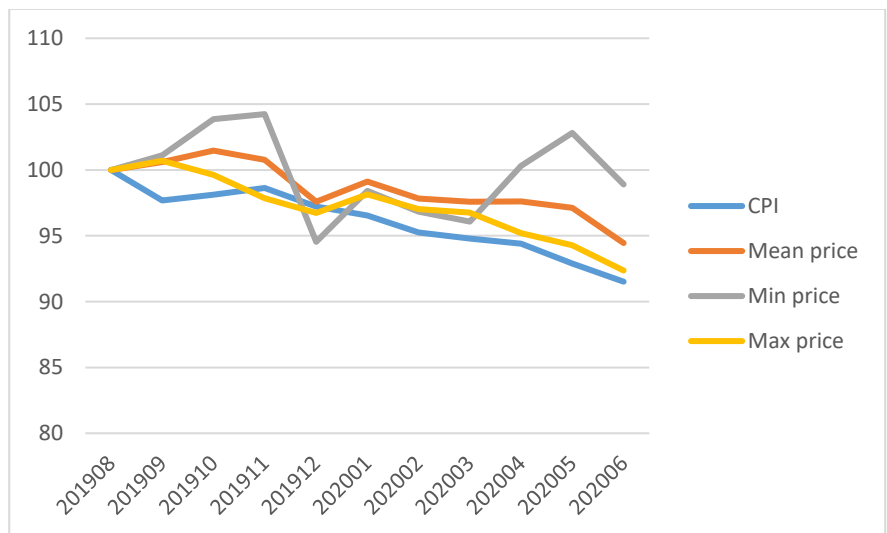
I figur 7 och 8 framkommer det att butiksurvalen påverkar indexutfallen annorlunda, där datorer verkar ha en lägre prisutveckling vid ett mindre urval medan det på sikt inte går att utskilja något sådant samband för mobiltelefoner.

Prismätning

Standard i denna studie har varit att aggregera det aritmetiska medelpriset över månaden, i varje butik eller över varje produkt. Som alternativ har det beräknats även min- och maxpriset per produkt och månad.



Figur 9: Månadskedjning av mobiltelefoner vid olika prisdefinitioner (augusti 2019=100)



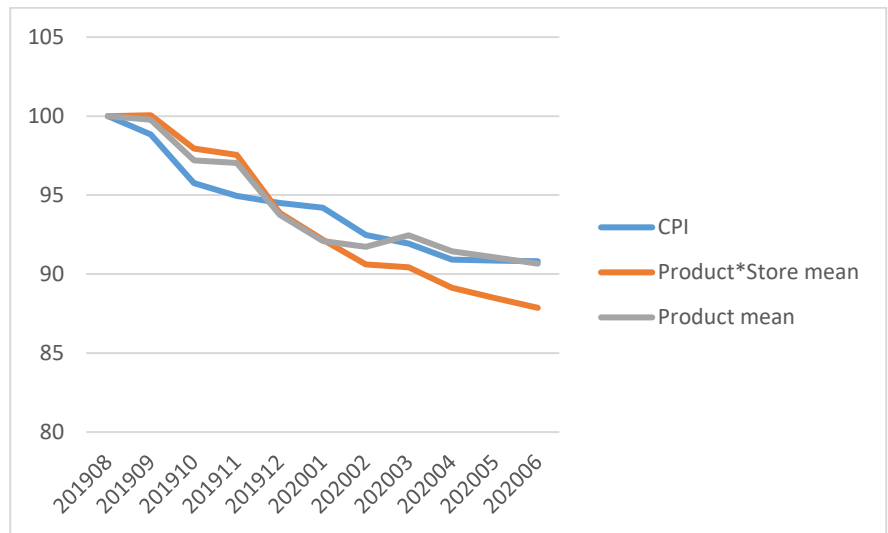
Figur 10: Månadskedjning av datorer vid olika prisdefinitioner (augusti 2019=100)

För mobiltelefoner följer indexutvecklingen för medelpriset KPI över tid, medan det högsta priset har en högre prisutveckling och det lägsta priset har en lägre prisutveckling än medelpriset. För datorer är sambanden mer svårtydliga.

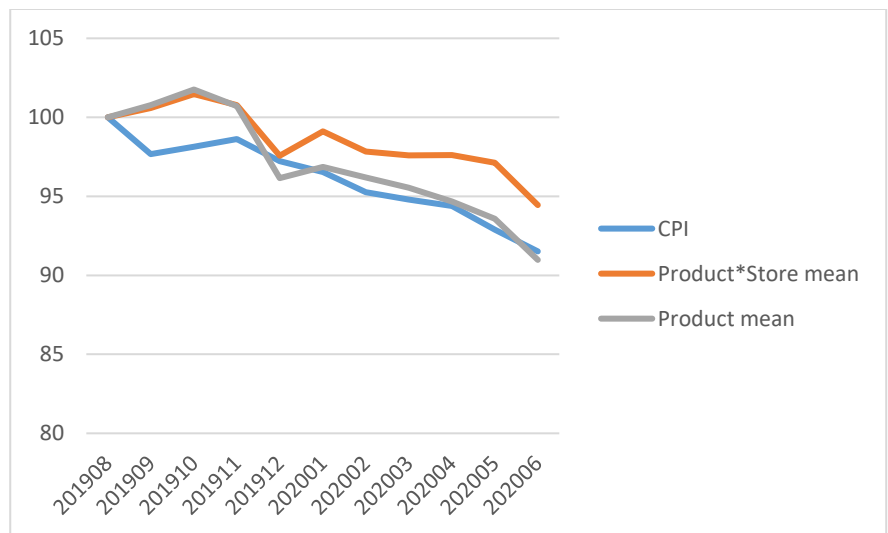
Produktdefinition

I studien har det även testats två olika definitioner av en prisnotering. Antingen definieras en prisnotering som det genomsnittliga priset på en produkt i en butik under en månad, eller som det genomsnittliga priset för en produkt under en månad. Då försäljningskvantiteter eller företagsvikter inte är tillgängligt, är aggregaten oviktade. För en prisnotering per butik aggregeras alla observationer

under samma månad aritmetiskt, för att senare aggregera priskvoterna för varje butik och produkt geometriskt. För en prisnotering definierad som en produkt, har alla priser noterade i alla butiker under månaden aggregerats aritmetiskt, för att sedan aggregera priskvoterna per produkt geometriskt.



Figur 11: Månadskedjning av mobiltelefoner vid olika definitioner av en prisnotering (augusti 2019=100)



Figur 12: Månadskedjning av datorer vid olika definitioner av en prisnotering (augusti 2019=100)

I figur 11 och 12 framkommer ett motsägande förhållande till produktdefinitionen mellan datorer och mobiltelefoner. Eftersom det inte har använts viktunderlag vid beräkningarna, blir den implicita vikningen av antalet butiker som erbjuder en produkt det mest approximativa alternativet till ett viktat index. Detta gäller givet att det finns ett samband mellan antal sålda och

spridning av butiker som säljer produkten. Den produktdefinition som tillåter implicit viktning av antalet butiker är definition om *produkt*butik*. För produktdefinitionen *produkt* väger alla produkter lika oavsett popularitet och oavsett storleken på utbudet i butiker.

Sammanfattning

I studien om indexberäkningar för mobiltelefoner och datorer framkommer två tydliga mönster:

- Indexnivå för månadskedjning av mobiltelefoner korrelerar starkt med storleken på antalet mest populära produkter i korgen
- Index med månadskedjning av mobiltelefoner och datorer påverkas olika mycket samt påverkas åt olika håll givet olika parametrar i undersökningsdesignen

Frågor till nämnden

Finns det anledning att fortsätta utvärdera olika delar av undersökningsdesignen på detta kvantitativa sätt genom att beräkna multipla alternativa index?