

Förutsättningar för nya kassaregisterdata

Nya transaktionsdata inom vissa produktgrupper ger potential att prismäta konsumtionen på ett helt annat sätt än genom insamling av enskilda priser. Dock medför de omfattande datamängderna andra utmaningar för att uppnå en relevant prismätning i nuvarande kontext med fast korg. Promemorian belyser vissa avvägningar som behövs och lyfter frågan om lämplig ambitionsnivå till nämnden för konsumentprisindex.

Bakgrund

För beräkning av konsumentprisindex har SCB möjligheten att använda transaktionsdata i KPI inom varugrupper så som möbler, hemtextilier, kläder, skor, byggvaror och andra diverse varor. Det handlar om nya dataunderlag utöver de varugrupper som tidigare har diskuterats i nämnden, så som exempelvis dagligvaror, alkohol, resor/resetjänster, hemelektronik och läkemedel.

I vissa varugrupper har dessa transaktionsdata redan implementerats, ibland jämte befintliga datakällor (manuell internetinsamling, webbskrapning) medan det ännu återstår för andra varugrupper. En väsentlig fråga som aktualiserats alltmer i takt med tillkomsten av ytterligare datakällor (transaktionsdata) är den praktiska användningen av denna typ av stora mängder data i nuvarande kontext med fastkorgsprincipen. Som också redovisas i Statistikens Framställning Konsumentprisindex (SCB 2025) förekommer vanligen urval med byten vid behov för att upprätthålla fastkorgsprincipen under pågående urvalsvår.

Syfte

I denna promemoria belyses en specifik problematik i användandet av större datamängder från enskilda butiker/kedjor, nämligen potentialen till stora urval kontra motsvarande upprätthållande av korgens storlek vid betydande oregelbundenhet/variation i sortimenten under året som prismätningarna görs.

Problematisering

En förekommande utmaning i hanteringen av transaktionsdata för konsumentprisindex är det omfattande behovet av kodning, alltså mappning till COICOP, vanligen från butikens/kedjans egen kategorisering, kvalitetskontroller av data och därefter ramframställan för (eventuella) urval till prismätningar. Denna process kan vara särskilt utmanande när transaktionsdata täcker produkter/grupper av produkter som har stora och varierande sortiment och även där viss och betydande försäljning inte sker regelbundet. En liknande fråga har tidigare lyfts i en promemoria om användningen av transaktionsdata för kläder och skor (Öhman och Tongur 2024), där det ofta förekommer ganska stora sortiment av snarlika produkter, ofta med många varianter, och säsongsvapor.

Ramframställan och kriterier

Vid tillgång till transaktionsdata för prismätningar uppkommer vanligen en fråga om ramavgränsning, nära knuten till urvalsarbetet inför prismätningar i det nya året, nämligen hur/vilka instegskriterier som ska tillämpas för att en produkt ska kunna ingå i urvalsramen (som alltså avser en förfluten tidsperiod) med avsikten att prismäta den under det kommande året.

Exempelvis kan ett kriterium vara hur många månader en produkt ska ha funnits i anslutning till basperioden för att den ska kvalificera in i ramen, eller därutöver hur pass mycket omsättning (relativ/absolut) produkten ska ha haft under en viss period (i ramen). Därtill kan det beaktas huruvida någon form av omsättningskriterium ska tillämpas löpande, alltså under mätperioden för att uppnå en viss måltäckning av försäljningen i butiken/kedjan. Ett annat typiskt förekommande kriterium är att försäljning av produkten ska förekomma i basperioden för att en priskvot ska kunna beräknas ($\text{priskvot} = \frac{\text{priset i referensmånaden}}{\text{priset i jämförelsetidpunkten}}$). Därtill kan det kravet kvalificeras ytterligare genom viss väsentlighet i termer av *antal* sålda enheter i basperioden.

Som kan förstås är dessa typer av kategoriska kriterier, framför allt för vissa sällanköpsvaror, något svåra att definiera i förhand i och med att produkter inom en viss varugrupp kan ha varierande livslängder ”på hyllan” i butikerna och/eller varierande regelbundenhet i försäljningen, vilket försvårar kontinuerliga mätningar i många fall. En konsekvens av sådan icke-regelbundenhet blir därmed att ta ställning till *hur ofta* och även *hur* motsvarande ersättningsbyten ska göras och om man bör ha en viss täckningsgrad i åtanke utifrån antingen den korg som formats i basmätningen eller den aktuella månadens försäljning.¹

Ur detta avseende är transaktionsdata helt annorlunda mot prisinsamling på internet (antingen manuell eller genom webbskrapning) där produkters priser vanligen finns så länge de ingår i sortimentet, oavsett om de säljs under månaden eller inte. I figur 1 illustreras fiktivt men baserat på transaktionsdata, hur en ram av produkter, fördelad över olika produktgrupper, varierar i omsättning (uttryckt relativt omsättningen i basen) under det efterföljande år som prismätningen görs.

Produkt-grupp	Bas+1 (jan.)	Bas+2 (feb.)	Bas+3 (mar.)	Bas+4 (apr.)	Bas+5 (maj)	Bas+6 (juni)
AA						
BB						
CC						
DD						
EE						

Figur 1 – Omsättning per månad jämfört med täckningen i basen

Färgsättningen går från mörkgrönt (en omsättning motsvarande basperiodens omsättning) till mörkrött (en omsättning mindre än en femtedel av basperiodens omsättning). Efter en viss tid försvinner vissa produkter helt medan vissa ”kommer och går”, vilket innebär att även basperiodens omsättning avspeglar en slags ”slump” i termer av vad som råkade sälja just då samtidigt som valet av annan basperiod hade resulterat i en delvis annan mängd för prismätningar.

I sammanhanget med fast korg skulle därför en urvalsansats innebära en löpande avvägning mellan att ha produkter som säljer (någorlunda) regelbundet och att ha en tillfredsställande täckning av konsumtionen i

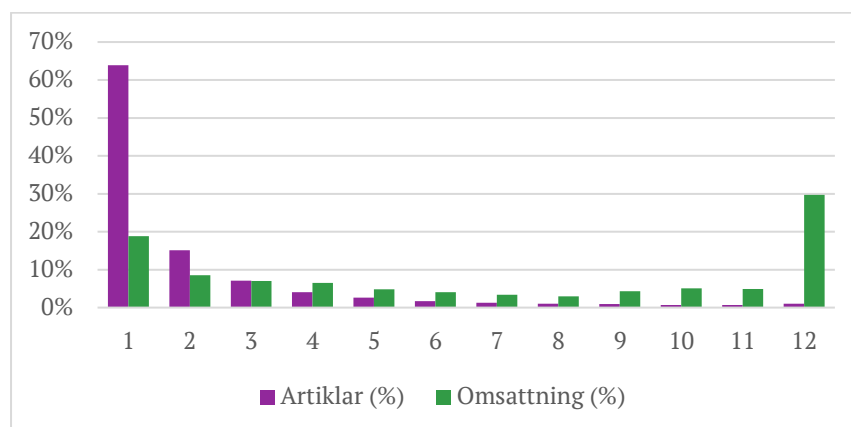
¹ Detta sistnämnda hanteras vanligen med så kallade *multilateral* metoder där en förutsättning är att transaktionsdata kontinuerligt hänförs till respektive produktgrupper i KPI-trädet och därutöver underliggande kategorier. SCB tillämpar inte denna typ av metoder ännu som dessutom inte är helt givna för den typer av produkter som berörs i promemorian.

termer av omsättning. Små urval gör det möjligt att upprätthålla korgen genom byten vid bortfall, vilket vanligen förutsätter kvalitetsvärdering av inbytta produkter för jämförbarhet, medan motsvarande process blir svårigen hanterbar i takt med att urvalen behöver göras större för att öka täckningen motsvarande respektive butiks/kedjas transaktionsdata.

En illustration av problemet

I det följande exemplifieras hur försäljning under ett år kan se ut för en sammansättning av sällanköpsvaror som kan ingå i en korg till KPI. Exemplet bygger på verkliga transaktionsdata över flera närliggande produktgrupper; i figur 2 illustreras de två parametrarna artikelförekomst och omsättning, fördelade över antal unika månader under det kalenderår som ramen täcker, och utan villkor att produkterna ska finnas i december (som är basperioden).

Som kan ses från de lila staplarna har ungefär 2/3 av sortimentet en livslängd på endast en månad (stapel #1) medan endast en hundradel (cirka 1%) av sortimentet säljer under samtliga 12 månader (stapel #12). Samtidigt framgår ur figuren genom de gröna staplarna att sortimentet som sålt under samtliga 12 månader står för 30% av omsättningen. Denna dimensionering har stor betydelse för den nuvarande och väl etablerade urvalsprincipen som baseras på försäljning närmast föregående år/åren liksom förekomst i basperioden (december i det närmast föregående året). Samtidigt bör inses att det inte är möjligt att i förväg veta livslängden nästkommande år för de produkter som ingår i urvalet för prismätningar (oberoende av urvalsmetod), vilket innebär att figur 2 endast ger en uppfattning ex post av potentialen för urvalet.



Figur 2. Fördelning av artiklar och omsättning över antal månader

En naturlig ambition i urvalsarbetet vid stora mängder transaktionsdata och utifrån de nämnda förutsättningarna vore att eftersträva en hög täckning under året med förhållandevis få produktbyten, väl avvägt för att hålla korgen representativ och samtidigt minimera handhavandet.

De ovan nämnda avvägningarna innebär således praktiska utmaningar i ramframställan och det följande urvalsarbetet, vilket kan illustreras genom följande tre olika ansatser för att bestämma en användbar ram för prismätningar under det efterföljande året y , utifrån de nämnda förutsättningarna som illustrerats i figur 2, som därmed betingas på förekomst under närmast föregående år $y-1$:

1. I det första fallet (definition 1) ansätts endast de produkter som närmast föregående år funnits samtliga tolv månader, motsvarande exemplifieringen i figur 2, stapeln längst till höger. Dessa utgör cirka en procent av antalet artiklar men närmare 30 procent av omsättningen.
2. I det andra fallet (definition 2) tillämpas likväl villkor på förekomst i december (som implicit krav eftersom december utgör basperiod för priskvoterna) men ett mildare krav på förekomst under minst sex månader (inte nödvändigtvis sammanhängande) under närmast föregående år. Dessa utgör cirka fem procent av antalet artiklar och står för ungefär 45 procent av omsättningen.
3. I det tredje fallet (definition 3) tillämpas endast instegsvillkoret förekomst i basperioden. Dessa 23 procent av produkterna behäftas därmed med ett visst (betydande) inslag av slump då det ej är villkorat på förekomst i andra månader närmast föregående år. Dessa produkter står för ungefär 60 procent av omsättningen.

I tabell 1 sammanfattas de tre fallen.

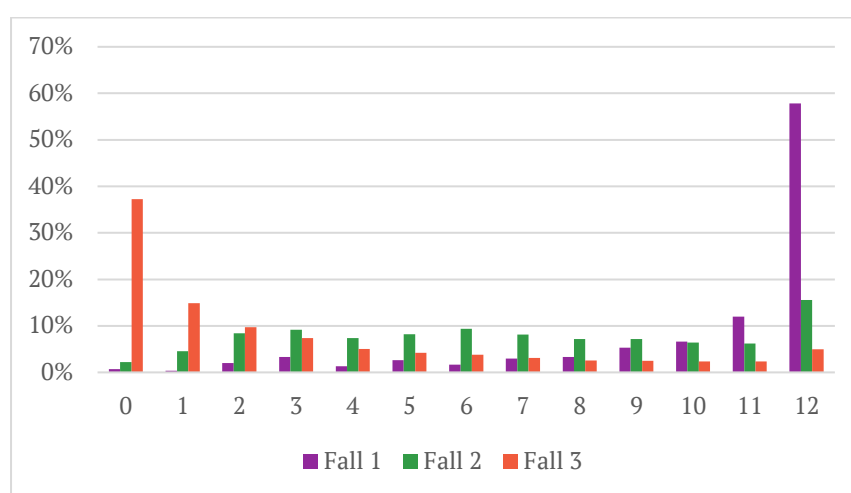
Tabell 1. Tre ytterligheter avseende täckning utifrån ramdefinition

	Fall 1	Fall 2	Fall 3
Andel produkter	1,01%	5,21%	23,64%
Andel omsättning	29,73%	45,75%	61,35%

Definition av ramen skulle utifrån definitionerna ovan innebära olika förutsättningar för korgen under det år som prismätningarna görs i

termer av handhavande av eventuella byten och beroende på eventuell urvalsmetod.

I figur 3 nedan visas på vad valet av ram skulle innebära i termer av relativ korgminskning om prismätningarna grundades på en korg enligt dessa tre definitioner av ramen, alltså under förutsättning att inga ersättningsbyten görs för att upprätthålla den ursprungliga täckningen. Som synes skulle det innebära helt olika relativ korgminskning under året som prismätningar görs.



Figur 3. Kvarvarande andel produkter antal månader efter basmånaden

Det första fallet enligt ovan illustreras genom de lila staplarna i figur 3. Efter tolv månader finns mer än halva mängden kvar (av den dryga procent av produkterna för helåret innan som ingår i prismätningarna) medan för det tredje fallet har mängden decimerats kraftigt till att innehålla knappt fem procent av den ursprungliga mängden (vilka tillsammans utgjorde nästan en fjärdedel av produkterna för helåret innan), i enlighet med parametrarna i tabell 1. En väsentlig aspekt i figur 3 är staplarna till vänster; en betydande mängd produkter är kortlivade med noll, en eller ett par månaders förekomst under kalenderåret (må vara oregelbundna försäljningar, tillfälliga sortiment och/eller säsonsprodukter). Dessa aspekter kan bli påtagliga vid täckningsmässigt större urval som i tredje fallet.

Avvägningar vid hantering av problemet

Prismätningar genom transaktionsdata innebär som nämnt en betydande insats för kodning och även kvalitetskontroll genom exempelvis granskning av data samt, som ovan nämnt, en insats för att identifiera en ram för prismätningar. I diskussionerna hittills har vi inte berört urvalsförfaranden utan enbart visat på hur olika typer av instegsvillkor påverkar definitionen av ramen för vidare prismätningar, där i sin tur urval kan tillämpas.

Ansatsen med urval är mycket vanlig även för transaktionsdata, särskilt där det finns viss beständighet i sortimenten, men i promemorian kontext är datamängderna vanligtvis mycket stora redan vid enstaka procents täckning av antal produkter i jämförelse med urvalen för traditionell insamling. En urvalsansats för att löpande upprätthålla en viss täckning skulle därför kunna innebära en mycket omfattande, egentligen helt ohanterlig, mängd byten som vanligen behöver handhas manuellt.² Exempelvis nämns i promemorian av Ribe (2017) att olika kvalitetsvärderingsmetoder är förenade med osäkerhet och risk för bias. I fallet med bedömningsmässiga kvalitetsvärderingar som är tillämpliga i sammanhanget nämns i promemorian *”Risk för bias särskilt om det finns en tendens att för det mesta överskatta kvalitetsförbättringar, eller att utelämna svåra kvalitetsbedömningar”*, vilket kan behöva betänkas i förhållande till mängden kvalitetsvärderingar som månatligen kan behöva göras för att upprätthålla korgen.

I termer av bias kan även följande aspekter beaktas:

- En begränsad mängd produkter utifrån första ramdefinitionen innebär en risk för bias genom undertäckning, särskilt av mer eller mindre tillfälliga eller kortlivade sortiment, så som vänstra delen av figur 3 avspeglar.
- En alltför omfattande mängd produkter som prismäts utifrån tredje ramdefinitionen (eller genom urval baserat på den mängden) kan innebära ett extremt antal månatliga handhavanden genom ersättningsbyten, medan det omvända att inte göra ersättningsbyten för att upprätthålla korgen, skulle innebära en potentiell

² Detta bör tolkas i kontexten att SCB har tillgång till många datakällor för prismätningar av den typ av produkter som berörs i promemorian, vilka tillsammans månatligen skulle innebära mycket omfattande mängd produkter vars ersättningsbyten måste hanteras, med given fastkorgsansats.

täckningsbias utifrån produkters livslängder som visats i figur 3, utifall att kortlivade produkter skulle ha en annan prisutveckling.

Breddad användning

I nära anslutning till promemorians frågeställningar är de fall där transaktionsdata även täcker försäljning av produkter som inte direkt betänktes och/eller ligger utanför den/de typiska produktgrupperna som avsågs vid valet av butik (eller kedja). I de fallen kan finnas en breddad (sam-) användbarhet av respektive transaktionsdatamängd men ställer då ytterligare krav för att identifiera en relevant sammanvägning av underlagen om dessa kommer från exempelvis helt olika branscher. Denna fråga har inte avhandlats i promemorian men tangerar urvalsfrågan i viss utsträckning i termer av sammanvägning av mängder med olika ”tyngd” enligt ramen och mängdernas representativitet gentemot varandra under året.

Frågor till nämnden

I och med den ökade tillgången till transaktionsdata inom flera olika områden skulle det vara till nytta med vägledande principer kring de avvägningar som behöver göras som grund för ramdefinitioner, urvalsarbete och manuellt handhavande genom ersättningsbyten. Nämnden inbjuds komma med synpunkter på det fortsatta arbetet och vilken ambitionsnivå som kan vara relevant vid införandet av nya transaktionsdata.

Referenser

Tongur, C. & Öhman, S. (2024). *Prismätningar för kläder och skor*. Stockholm: PM till nämnden för KPI, sammanträde nr 18 (maj 2024), SCB.
<https://www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/18/7.-forutsattningar-for-prisinsamling-av-coicop-03.pdf> [Hämtad 2025-10-14]

Statistiska centralbyrån (2025). *Statistikens Framställning Konsumentprisindex*.
https://www.scb.se/contentassets/a1e257bb3a574420b9d3f2ff59851c0a/pr0101_staf_2025.pdf [Hämtad 2025-10-14]

Ribe, M. (2017). *Bias i KPI*. Stockholm: PM till nämnden för KPI sammanträde nr 3 (oktober 2017), SCB.
<https://www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/3/3-2-bias-i-kpi.pdf> (Hämtad 2025-10-14)