

Kassaregisterdata för dagligvaror – historik och planer framåt *

I denna PM ges en kortare uppdatering kring SCB:s arbete med kassaregisterdata för dagligvaror. Nämnden välkomnas att komma med synpunkter inför det fortsatta utvecklingsarbetet.



Medfinansieras av
Europeiska unionen

Bakgrund

SCB har använt kassaregisterdata (KRD) för dagligvaror i produktionen av Konsumentprisindex och relaterade inflationsmått sedan 2012. I princip innebär den implementering som gjorts att datamaterialet inlemmats i befintlig beräkningsstruktur, vilken ursprungligen skapades för traditionell (intervjuarinsamlad) data. Såväl tekniskt upplägg som metodval har med andra ord anpassats efter de förutsättningar som rådde sedan tidigare.

Under senare år har emellertid ett antal metodstudier på området genomförts av SCB. Automatiserade lösningar för extremvärdes- och byteshantering har testats, såväl som månadskedjning och multi-lateral indexmetoder. Resultat från studierna har dock ännu inte tagits in i produktionen.

I denna PM ges en kortare uppdatering av SCB:s planer vad gäller användningen av kassaregisterdata för dagligvaror i KPI och relaterade mått. För att placera frågan i ett sammanhang görs även en historisk tillbakablick. Nämndens ledamöter inbjuds att notera redovisningen samt att komma med eventuella synpunkter inför det fortsatta arbetet.

* Det utvecklingsarbete som beskrivs har delvis finansierats med medel från Eurostat. Åsikter och synpunkter i denna promemoria återspeglar dock inte nödvändigtvis EU:s eller Eurostats ståndpunkter.

Tidig historik (2009–2019)

Kassaregisterdata för dagligvaror har diskuterats i KPI-nämnden åtskilliga gånger. Det första nämndunderlag som skrevs på området förefaller ha varit Friberg och Norberg (2009). Här kan man ana en försiktig inställning till ämnet; bland annat konstateras att det ”återstår en hel del problem att belysa och eventuellt lösa innan [kassaregisterdata] via EAN-koder kan ersätta prisnoteringar som insamlingsmetod för KPI”. I protokollet står vidare att SCB:s ambition *inte* är att genomföra ”någon stor utökning av urvalet” utan enbart att ”använda [datamaterialet] som en alternativ metod för datainsamling”.

Åren därpå bedrivs mycket utvecklingsarbete på SCB inom detta område, och ämnet förekommer därmed också en hel del i nämndens diskussioner. Bland annat redovisar SCB en jämförande studie där KRD ställs mot manuellt insamlade uppgifter (Norberg m.fl., 2011), vilken visar på små skillnader mellan insamlingssätten.¹ Därutöver diskuteras hanteringen av rabatter och kampanjer i kassaregistermaterialet samt hur aggregering av veckovisa uppgifter till månadspriser bäst ska göras.² Från protokollet kan utläsas att flera av KPI-nämndens ledamöter anser att SCB visserligen bör arbeta för att inkorporera KRD-materialet i beräkningarna, men att man då bör fokusera på att efterlikna befintlig design.³ (När det kommer till hanteringen av rabatter är dock hållningen att sådana bör inkluderas, ”med undantag av ’bäst före’-rabatter”.⁴)

Vidare diskuterar nämnden risken för nedåtgående bias om ut- och ingående artiklar (EAN-koder) inte länkas korrekt för kassaregistermaterialet. Som svar på detta anför SCB att principerna som används är desamma som för intervjuarinsamlade data, och att ett ”mycket viktigt steg i granskningen av kassaregisterdata är att göra en marknadsanalys och då grundligt undersöka varför en EAN-kod försvunnit från marknaden”. Man konstaterar vidare att det ur denna aspekt är ”bra med urvalsbaserad prisinsamling så att man med realistisk resursinsats kan följa priset över varubyten” och att ”risken för bias annars är stor”.⁵

¹ I analysen används listpriser från KRD-materialet, inte transaktionspriser.

² Norberg m.fl. (2012), Sammar (2011; 2012a; 2012b), Sammar och Norberg (2012), Sammar m.fl. (2013a; 2013b), Sammar och Ribe (2012) och Tongur m.fl. (2012).

³ Vad gäller frågan om veckovis aggregering förordas först geometrisk formel, vilket dock under 2012 ändras till kvantitetsvägd aritmetisk sammanvägning. (Se protokollen för sammanträden nr 244 och 246.)

⁴ Se protokollet för sammanträde nr 245.

⁵ Se protokollet för sammanträde nr 248.

En annan fråga som tar mycket plats i bland annat nämndens diskussioner, i synnerhet från 2016 och framåt, är behovet av att utveckla särskilda metoder för att kunna nyttja KRD även för prismätning av färskvaror.⁶ Här är det framförallt valet av produkt-specifikation som diskuteras, d.v.s. över vilken mängd genomsnittspriser ska beräknas.⁷ Till en början väljer SCB att räkna priset per produkt för dessa varor genom att ta ett enhetsvärde över liknande artiklar, men detta ändras senare till att omfatta enbart den ”mest sålda” varan. Argumentet för ändringen är att man vill räkna på ett likartat sätt för alla datakällor, och att ”mest-sålda”-förfarandet mest liknar intervjuarsituationen. (Vid denna tidpunkt görs fortfarande en del av prisinsamlingen av intervjuare.) I SCB:s underlag till nämnden anges dock även att detta val kan komma att justeras i framtiden.⁸

Från 2016 börjar också frågan om val av indexformel få ökat fokus i KPI-nämnden. Från början diskuteras främst möjligheten att byta metod till någon form av månadskedjning (”dynamisk ansats”), men snart kommer även multilaterala indexmetoder upp på dagordningen.⁹ SCB genomför ett par mindre jämförande studier på området, men dessa baseras i regel på mindre datamängder (exempelvis en delmängd av produktsortimentet och/eller en kortare tidsperiod), då den befintliga datastrukturen försvårar mer genomgripande analyser. (Det främsta problemet är att det fullständiga datamaterialet ännu inte produktgruppsklassificerats).¹⁰ Följaktligen är det heller ingen av dessa metoder som implementeras i produktionen. Mer omfattande analyser har dock gjorts inom ramen för de grantsprojekt som drivits under senare år, vilka beskrivs i mer detalj i nästa avsnitt.

Figur 1 sammanfattar tidsplanen för införandet av KRD för dagligvaror i KPI-beräkningarna. Inför 2012 ersätter SCB en del av prisinsamlingen för en första butikskedja med kassaregisterdata, vilket året därpå utökas till att omfatta ytterligare två kedjor. Inför 2017 införs KRD även för färskvaror för en av de tre kedjorna, och 2018 tas även de andra två kedjornas data in. Från och med 2019 omfattas hela dagligvaruområdet av KRD. Sedan dess har det beräkningsmässiga upplägget varit

⁶ Tongur och Sandén (2016), Bilius och Tongur (2017a), Bilius m.fl. (2017).

⁷ Med ”produktspecifikation” menar vi den operationella definitionen av vad som utgör en individuell produkt. Det är på denna nivå ett pris (enhetsvärde) och en eventuell kvantitet beräknas. (Se också figur 1 i Ståhl, 2021.)

⁸ Bilius m.fl. (2017).

⁹ Månadskedjning förordas av flera ledamöter under sammanträde nr 1, och multilaterala metoder bland annat under sammanträden nr 3 och 4.

¹⁰ Se Bilius och Tongur (2017b) och Bilius m.fl. (2018).

i princip oförändrat, så när som på en del anpassningar vilka krävts till följd av ändringar i dataleveransernas format och struktur.¹¹



Figur 1: Några centrala milstolpar i implementeringen av kassaregisterdata för dagligvaror i svenska KPI och relaterade inflationsmått.

EU-finansierade projekt (2021–2026)

Två delvis EU-finansierade projekt har bedrivits av SCB mellan 2021 och 2026. Dessa har varit relativt brett formulerade. Bland annat har de omfattat en översyn av aggregeringsstruktur och urvalsdesign, utvärderingar av dynamiska och multilaterala indexmetoder, samt försök att automatisera dagens manuella produktbytesansats (antingen genom gruppering av artiklar eller med hjälp av imputeringar).¹²

En viktig slutsats från det första av de två projekten var att ett stegvis upplägg för framtida utvecklingsarbete torde vara den mest framkomliga vägen; detta då olika metodmässiga aspekter är starkt beroende av varandra. (Exempelvis är det svårt att analysera val av indexformel utan att först ha klassificerat datamaterialet och beslutat om lämplig produktspecifikation.) Det upplägg som föreslogs av SCB 2023 följde denna stegvisa ansats.¹³

Som ett första steg noterades att datalagring och övrig IT-infrastruktur behöver uppdateras. Det historiska datamaterialet måste också produktgruppsklassificeras, samtidigt som en process för att varje månad klassificera nyttillkomna data måste komma på plats. Först efter

¹¹ Bland annat efterfrågar SCB numera i större utsträckning data för samtliga butiker inom respektive butikskedja; jmf Eliasson (2023).

¹² Projektsresultaten sammanfattas i två rapporter; Ståhl (2023a; 2026).

¹³ Se Ståhl (2023b).

det kan automatiserade metoder vilka ersätter (åtminstone stora delar av) dagens manuella varubyten implementeras. När allt detta är på plats, kan urvalen i produktledd utökas och metoder för att nyttja månatlig kvantitetsinformation utredas närmare med utgångspunkt i det fullständiga materialet.

Det andra projektet, vilket slutfördes i september 2026, följde ovanstående plan och fokuserades på de första delstegen. Det som utreddes var med andra ord främst möjligheterna att effektivisera och automatisera processerna för års- och månadsproduktionen. Rent praktisk innebar detta bland annat att den databasstruktur som skissats på i det första projektet förfinades, och ett särskilt editeringsverktyg utvecklades. Vidare genomfördes en empirisk studie där olika automatiska alternativ till de manuella länkningarna testades.

En slutsats från detta projekt var att det för de flesta produkter *inte* var lämpligt att gruppera materialet baserat på den kategori-indelning som finns tillgänglig, för att sedan använda dessa grupper som produktspecifikationer. Detta då sådana grupper tenderar att bli alltför breda. Mer bias (i form av s.k. ”enhetsvärdesbias”) riskerar med andra ord att tillföras än den bias som eventuellt undviks (vilken istället beror på missade *relaunches* och ”krympflation”). Däremot indikerade resultaten att automatiserad länkning (gruppering) av artiklar baserat på deras namn skulle kunna vara en framkomlig väg, och att information om exempelvis märke och kategori kan vara till hjälp för att hitta dessa matchningar.

Diskussion

När KRD för dagligvaror implementerades i KPI-beräkningarna valde SCB ansatsen att efterlikna den metod som sedan tidigare användes för intervjuinsamlade data. För detta angav Norberg m.fl. (2011) två huvudsakliga skäl. Det första var att SCB inte skulle mäka med att månatligen klassificera all inkommande data till rätt produktgrupp. Det andra var att det fanns behov av att manuellt länka utgående artiklar med nya varianter, något som bedömdes bli alltför tidskrävande om hela datamängden skulle hanteras.¹⁴ Utöver dessa två skäl kan även en

¹⁴ Om metoden ”compute index from a census based on all products for which scanner data are available” står bland annat: ”SCB has made assumptions that there might be some problems using this method. SCB is not convinced that NSI officers have enough knowledge to classify over ten thousand products into Coicop-groups with sufficient quality. [...] Finally substitutes cannot be handled automatically, e.g., when the package size is altered (for example the number of napkins decrease in a package), the EAN-code is altered and the price remains unchanged then the implicit price increase will not be calculated but linked to show no change if not processed manually.” Ett tredje skäl som anges i artikeln är svårigheterna i att justera för pant (för exempelvis läsk), om hela materialet skulle

mer praktisk aspekt utläsas ur samtida nämndprotokoll, nämligen bristen på ett flexibelt systemstöd vilket även klarar av att hantera de större datamängderna. Vi kan dock konstatera att, inte minst med tanke på den tekniska utveckling som skett under de senaste 15 åren, inget av dessa argument fortfarande torde vara giltigt idag.

Utifrån ovanstående resonemang kan det tyckas givet att SCB bör uppdatera beräkningsmetoderna för dagligvaruområdet. Samtidigt är läget inte akut; betydande kvalitetsförbättringar har trots allt redan gjorts i och med införandet av kassaregisterdata. Inte minst har urvalsstorlekarna ökat och variansen därmed minskat. Dessutom torde dagens skattning av månatliga genomsnittspriser (som ett enhetsvärde över månadens försäljningar) vara en betydligt bättre representation av vad konsumenter faktiskt betalar än det enskilda hyllpris som intervjuarna kunde observera. Eftersom dagligvaruområdet inte är ett av KPI-korgens mest dynamiska, väntar vi oss heller inte en kraftig effekt av att börja använda månadsvikter. Med all detta sagt är dock prisutvecklingen för dagligvaror, och särskilt livsmedel, ofta under lupp, och har även en relativt stor vikt i KPI. SCB bör därmed fortsätta arbetet med att säkerställa att dessa delindex håller högsta möjliga kvalitet.

Planer framåt

Figur 2 beskriver SCB:s planer inom dagligvaruområdet. Inför 2027 är tanken att beräkningarna ska lyftas ut ur dagens produktions-system. En rad effektiviseringar vad gäller klassificering av data och rambearbetning är också planerade. Dessutom införs möjligheten att mäta en kombination av flera olika artiklar under samma "produktkod", för att förbättra mätningen men också för att förenkla anpassning till mer automatiserade bytesmetoder.

I övrigt planerar SCB under 2027 även att arbeta vidare med data-kvalitén. Särskilt variablerna "märke" (t.ex. *Marabou*), "storlek" (t.ex. *100*) och "enhet" (t.ex. *gram*) är viktiga, eftersom denna information antas användbar för att få till så bra artikelmatchningar som möjligt. *Machine learning* avses även testas som stöd för klassificeringen av data till elementära produktgrupper.

Inför 2028 planeras en ökning av urvalsstorlekarna. Beroende på de resultat som uppnåtts innan dess, vad gäller såväl klassificering som automatisering av länkningsmomentet, kan detta innebära en övergång

inkluderas; denna aspekt återstår att studera för berörda produktgrupper, och skulle fortfarande kunna utgöra skäl att hålla nere urvalsstorleken för just dessa artiklar.

till att mäta *hela* KRD-materialet, eller en större delmängd av det. Möjlig implementering av (direkt) superlativ eller multilateral ansats sker också tidigast 2028, då motsvarande analys väntas kunna påbörjas tidigast under våren 2027.

2026	Ramberbetning och urval inför 2027 görs helt med utgångspunkt i den kassaregisterdata som ligger till grund för KPI-produktionen
2027	<u>Inför januari:</u> Övergång till ny datastruktur och månadsvis klassificering av data Möjlighet till "kontrollerad gruppering" av artiklar införs <u>Under året:</u> Metoder för automatiska länknings införs och förfinas (men granskas också manuellt) Vidare granskning och editering av historiskt material (särskilt märke, storlek och enhet) ML-lösning testas för produktgruppsklassificering Studier av superlativ eller multilateral ansats
2028	Ytterligare implementering av automatiska länknings i produktionen Ytterligare effektivisering av klassificerings- och editeringsarbetet Utökning av urvalet Eventuell implementering av superlativ eller multilateral ansats

Figur 2: Uppdaterad tidsplan för vidareutvecklingen inom KRD för dagligvaror.

Referenser

- Bilius, Å., Bubuioc, R. och Tongur, C. (2017). *Bestämning av prisvariabeln vid utökad användning av kassaregisterdata för viktvaror*. [PM till nämnden för Konsumentprisindex, möte nr 3](#).
- Bilius, Å., Ståhl, O. och Tongur, C. (2018). *Alternativa beräkningsmetoder för kassaregisterdata: kaffe-index*. [PM till nämnden för Konsumentprisindex, möte nr 4](#).
- Bilius, Å. och Tongur, C. (2017a). *Vägningstal för färsk frukt och färska grönsaker från kassaregisterdata*. [PM till nämnden för Konsumentprisindex, möte nr 2](#).
- Bilius, Å. och Tongur, C. (2017b). *Empirisk jämförelse av fast varukorg visavi månadsmatchning för dagligvaror*. [PM till nämnden för Konsumentprisindex, möte nr 3](#).
- Eliasson, J. (2023). *Nya ambitioner för livsmedelsstatistiken*. [PM till nämnden för Konsumentprisindex, möte nr 17](#).
- Friberg, R. och Norberg, A. (2009). *Om användning av scannerdata för prisinsamlingen till KPI – lägesrapport i september 2009*. PM till nämnden för Konsumentprisindex, möte 240.

Nilsson, P. och Ståhl, O. (2019). *Towards a Roadmap for Efficient Use of Electronic Transaction Data in the Swedish CPI*. [Artikel skriven för Ottawagruppens 16:e möte, Rio de Janeiro, Brasilien, 8–10 maj 2019](#).

Norberg, A., Sammar, M. och Tongur, C. (2011). *A Study on Scanner Data in the Swedish Consumer Price Index*. [Artikel skriven för Ottawagruppens 12:e möte, Wellington, Nya Zeeland, 4–6 maj 2011](#).

Norberg, A., Sammar, M. och Tongur, C. (2012). *Scanner data – comparability issues*. [PM till nämnden för Konsumentprisindex, möte nr 245](#).

Sammar, M. (2011). *Kassaregisterdata i KPI*. PM till nämnden för Konsumentprisindex, möte nr 244.

Sammar, M. (2012a). *Rabatter i kassaregisterdata för dagligvaror*. [PM till nämnden för Konsumentprisindex, möte nr 246](#).

Sammar, M. (2012b). *Prisbegreppet i KPI*. [PM till nämnden för Konsumentprisindex, möte nr 246](#).

Sammar, M. och Norberg, A. (2012). *Sammanvägningsmetod över tre veckor för kassaregisterdata i KPI*. [PM till nämnden för Konsumentprisindex, möte nr 246](#).

Sammar, M., Norberg, A. och Tongur, C. (2013a). *Fortsatt arbete med kassaregisterdata 2013*. [PM till nämnden för Konsumentprisindex, möte nr 248](#).

Sammar, M., Norberg, A. och Tongur, C. (2013b). *Issues on the use of scanner data in the CPI*. [Artikel skriven för Ottawagruppens 13:e möte, Köpenhamn, Danmark, 1–3 maj 2013](#).

Sammar, M. och Ribe, M. (2012). *Kassaregisterdata i KPI – utökad tillämpning för dagligvaror*. [PM till nämnden för Konsumentprisindex, möte nr 247](#).

Ståhl, O. (2021). *Aggregeringsprinciper inom COICOP 01 och 02*. [PM till nämnden för Konsumentprisindex, möte nr 12](#).

Ståhl, O. (2023a). *Towards More Efficient Use of Transaction Data in the Swedish HICP – Eurostat grant project summary report*. [Projektrapport](#).

Ståhl, O. (2023b). *Metodutredning för COICOP 01 och 02 – Projekt-rapport*. [PM till nämnden för Konsumentprisindex, möte nr 16](#).

Ståhl, O. (2026). *Further modernization of the Swedish HICP – work package 2 report*. [Projektrapport](#).

Tongur, C., Norberg, A. och Sammar, M. (2012). *Prisindex med och utan rabatter*. [PM till nämnden för Konsumentprisindex, möte nr 246](#).

Tongur, C. och Sandén, B. (2016). *Viktvaror från kassaregisterdata*. [PM till nämnden för Konsumentprisindex, möte nr 1](#).