

ES/PR  
Åsa Bilius, Can Tongur

# Empirisk jämförelse av fast varukorg visavi månadsmatchning för dagligvaror

## *För information/diskussion*

Detta är en empirisk studie av förhållandet mellan månadsmatchning visavi den befintliga fasta varukorgen. Tillämpningen görs på de kassaregisterdata SCB får för dagligvaror (COICOP 01) och är ett första steg i att utreda vilken indexformel som kan vara lämpligast för kassaregisterdata. Nämnden bjuds in till diskussion om hur SCB bör beakta frågan i fortsättningen.

## INNEHÅLL

1	Bakgrund.....	2
2	Syfte.....	2
3	Indexformulering och förutsättningar.....	2
3.1	Månadsmatchningsprincip.....	3
3.1.1	Urvalskriterium: tröskelvärde som instegsvillkor.....	3
3.1.2	Indexformulering.....	3
3.1.3	Automatisk/mekanisk bestämning av produktgrupp.....	3
3.2	Fastkorgsindex.....	4
3.2.1	Indexformulering.....	5
3.3	Variant: kontrollerad månadsmatchning.....	5
4	Resultat.....	5
5	Diskussion och fortsatta studier.....	7
6	Referenser.....	8
7	Appendix.....	8
7.1	Tröskelvärdesbestämning.....	8
7.2	Aggregering över uppgiftslämnare.....	8



## 1 Bakgrund

I användningen av kassaregisterdata avviker Sverige från andra europeiska länder på så sätt att fastkorgprincipen beaktas strikt genom ett årligen fast urval av produkter. I andra länder beaktas en mer dynamisk urvalsmekanism och därigenom utnyttjas större mängder data. Nederländerna och Norge tillämpar en månadsmatchningsprincip av produkter som finns i bägge av två intilliggande månader.

För KPI:s prismätningar i dagligvaruhandeln används en fast varukorg med direkt jämförelse av varje månads priser mot respektive pris i basen, december föregående år. Indexformel är Jevons index på den lägsta indexnivån, s.k. elementärt aggregat. För dagligvaror är de två befintliga datakällorna kassaregisterdata (från de största dagligvaruhandelskedjorna) och manuell prisinsamling från övriga dagligvaruhandeln. Både för kassaregisterdata och manuell prisinsamling används ett begränsat urval av produkter och butiker. I urvalsdragningen, som sker årligen, används omsättningsinformation (i kronor) för året närmast före basperioden, vilket blir två kalenderår före mätmånaderna. Dock uppdateras det dragna urvalet med avseende på senast identifierade EAN-kod, i.e. en aktualitetskontroll görs mot aktuellt kassaregisterdata precis innan EAN-koden (produkten) tas med i KPI:s basmånad december.

## 2 Syfte

Syftet med denna studie är att ta ett första steg i att jämföra det befintliga förfarandet i KPI, fastkorgsindex med begränsat urval, mot en dynamisk ansats med större urval, s.k. *månadsmatchning*.

Månadsmatchning, s.k. *matched items* används av exempelvis den nederländska statistikbyrån (CBS) för deras kassaregisterbaserade dagligvarusystem och avsikten häri är att illustrera eventuella empiriska skillnader mellan dessa två ansatser under förenklade omständigheter.

SCB planerar fortsatta analyser och eventuellt en översyn av den befintliga hanteringen av kassaregisterdata i KPI.

## 3 Indexformulering och förutsättningar

I det följande redogörs för två principiellt olika indexformuleringar för elementär aggregatet.

1. Månadsmatchningsprincip genom automatiska urval, vilket innebär jämförelse av pris mellan parvisa månader ( $m, m-1$ ).
2. Fastkorgsindex för jämförelse av pris i aktuell månad  $m$  med basperioden 0 i december föregående år ( $y-1$ ).

Därutöver skapas ett slags kontrollerad benchmark för månadsmatchningen för att kunna identifiera eventuella effekter av det automatiska urvalet.

Aggregeringen över uppgiftslämnare förklaras i appendix 7.2.

### 3.1 Månadsmatchningsprincip

Detta förfarande bygger på att de EAN-koder som finns i datamaterialet i bägge av två intilliggande månader tillåts ingå i indexberäkningen. Urvalet regleras av ett tröskelvärde för omsättning och genom automatisk matchning enligt redogörelse nedan. Index skapas som utvecklingen mellan parvisa månader ( $m, m-1$ ) och kedjas till basperioden 0 (december föregående år). Urvalsmetoden av produkter (EAN-koder) är automatisk/mekanisk identifiering, vilket är helt annorlunda mot det sannolikhetsurval som används idag för kassaregisterdata.

#### 3.1.1 Urvalskriterium: tröskelvärde som instegsvillkor

Ett tröskelvärde tillämpas för inkludering av varor i indexberäkningen. Tröskelvärdet avser försäljningen per EAN-kod uttryckt som en andel i produktgruppens totala försäljning, i detta fall för båda av de två ingående månaderna. I appendix formuleras detta. Tröskelvärdet kan bestämmas bedömningsmässigt. För denna studie har ställts ett minimikrav att försäljningsandelen i kronor ska överstiga en procentenhet i båda månaderna i den identifierade produktgruppen.

#### 3.1.2 Indexformulering

Månadsmatchningsprincipen innebär att varje månadslänk formuleras som en geometrisk summa (ovägd) av priskvoter enligt

$$\pi_a^{(y,m)/(y,m-1)} = \prod_{i=1}^{n_a^{(y,m),(y,m-1)}} \left( \frac{p_i^{y,m}}{p_i^{y,m-1}} \right)^{1/n_a^{(y,m),(y,m-1)}} \quad (1)$$

Månadslänkarna i (1) sätts ihop, kedjas, genom multiplikation till ett index med basperioden  $y_0, m_0$ :

$$P_a^{(y,m)/(y_0,m_0)} = P_a^{(y,m-1)/(y_0,m_0)} \cdot \pi_a^{(y,m)/(y,m-1)}, \quad (2)$$

där ekvation (2) är en rekursiv beräkning baserad på varje tidigare länkning inom året av kedjelänken i (1),  $P_a^{(y,m-1)/(y_0,m_0)}$ .

#### 3.1.3 Automatisk/mekanisk bestämning av produktgrupp

En förutsättning för den ansats som tillämpats i denna studie är automatisk/mekanisk identifiering av produktgrupper. Tillvägagångssättet har varit att ”baklänges-identifiera” produktgrupperna i KPI från kassaregisterdata genom att utnyttja KPI:s nyckeltabell som används för matchningen mellan EAN-koder och produktkoder. Nyckeltabellen är grunden i inläsningen av kassaregisterdata till KPI:s system för framställning av fastkorgsindexet och uppgår till ungefär 750 EAN-koder per kedja, av vilka en betydlig mängd är gemensamma mellan kedjorna. För denna studie har nyckeltabellen använts

”baklänges” enligt nedan matchningsprincip.

#### *Matchningsprincip*

Varje EAN-kod hör till en specifik varugrupp i dagligvarukedjornas kassaregisterdata. Dessa varugrupper är kedjornas egna benämningar. Om man tänker ”baklänges” kan en varugrupp i datamaterialet identifieras via de EAN-koder som används i KPI-urvalet. Eftersom varje EAN i urvalet är kopplad till en produktkod får man därmed produktgruppen enligt COICOP. I de fall där en varugrupp i data kopplar till flera produktgrupper i KPI har den först påträffade produktgruppen valts av förenklings-skäl. För de valda varugrupperna har sedan *samtliga* EAN-koder tillåtits ingå i urvalet, alltså inte enbart de EAN-koder som finns i KPI:s urval. Ansatsen exemplifieras i Tabell 1.

#### *Exempel på matchningsprincipen*

En EAN-kod för kanelbullar har i början av året länkats manuellt till produktgruppen ”1113 – Vetebröd” i nyckeltabellen, alltså KPI:s urval. Samma EAN-kod relaterar till varugruppen ”Fikabröd fryst” i indatafilen från en viss kedja. I det automatiska urvalet kommer alla EAN-koder som tillhör varugruppen *Fikabröd fryst* tilldelas produktgruppen 1113 – Vetebröd.

Tabell 1: Exempel på matchningsprincipen för produktgruppen 1113 – Vetebröd

NYCKELTABELL			INDATA KASSAREGISTER		UTDATA URVAL		
EAN-kod	Produktnamn	Prodgrupp	EAN-kod	Varugrupp	EAN-kod	Varugrupp	Prodgrupp
12345678	KANELBULLAR	1113	12345678	Fikabröd fryst	12345678	Fikabröd fryst	1113
12399999	GIFFLAR	1113	12388888	Fikabröd fryst	12388888	Fikabröd fryst	1113
			...		...		
			12399999	Fikabröd mjukt	12399999	Fikabröd mjukt	1113
			...				
			12365432	Fikabröd kakor			
			12311111	Fikabröd kakor			

Matchningsprincipen utgår från att vi antar att kedjornas varugrupsindelning överensstämmer någorlunda med den etablerade COICOP klassificeringen, vilket är ett förenklingsantagande för kunna genomföra denna studie. Denna tillämpning antas ändå vara principiellt analog med motsvarande implementeringar i andra länder, dock utan samma stringens i identifieringen av produktgrupper via varugrupper.

Effekten av denna mekaniska ansats kan vara att det för vissa produktgrupper blir ett ogynnsamt litet antal produkter (EAN-koder) som inkluderas beroende på att KPI-urvalet inte har någon koppling till varugrupperingarna i kassaregisterdata.

## 3.2 Fastkorgsindex

Som jämförelseindex används den befintliga indexkonstruktionen för dagligvaror. Detta är prisjämförelse mellan period  $m$  och basperioden  $0$ , december föregående år,  $y-1$ , med Jevons index. För denna mängd är urvalet

noggrant genomarbetat och månatligen uppdaterat för ersättningsprodukter och varje EAN-kod kopplar till en produktgrupp i KPI, oberoende av kedjornas egna varugrupperingar.

### 3.2.1 Indexformulering

Index beräknas per produktgrupp  $a$  enligt följande formulering av Jevons index:

$$P_a^{0,m} = \prod_{i=1}^{n_a} \left( \frac{P_i^m}{P_i^0} \right) \quad (3)$$

Det bör noteras att i (3) utnyttjas för denna studie inga enskilda vikter mellan produkter.

### 3.3 Variant: kontrollerad månadsmatchning

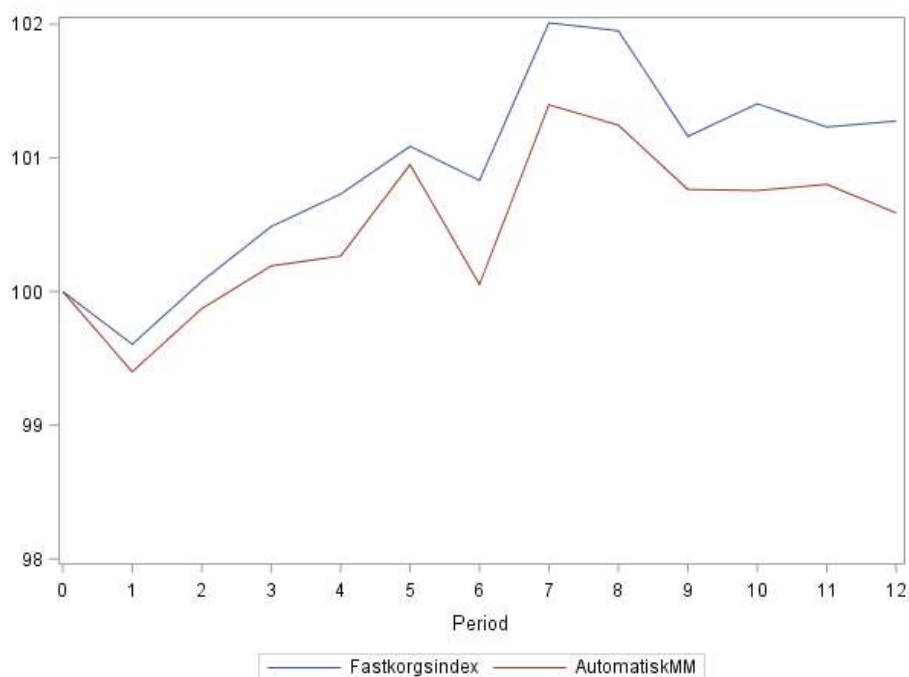
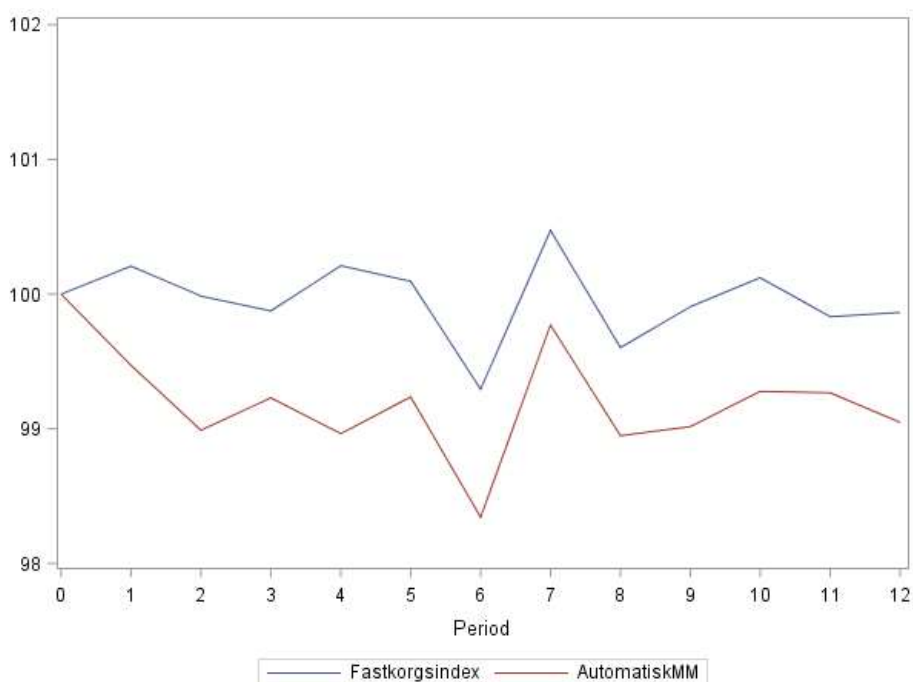
Som referensmetod för månadsmatchningsansatsen med automatiskt (=okontrollerat) urval tillämpas en månadsmatchning med kontrollerat urval. Det kontrollerade urvalet är baserat på det faktiska produkturvalet för fastkorgsindex men beräkning sker enligt månadsmatchningsprincipen i avsnitt 3.1. Det innebär att produktgruppstillhörigheten är säkerställd men urvalet blir sannolikt i många fall påtagligt mindre än vid okontrollerad månadsmatchning. Syftet med den kontrollerade metoden är att försöka identifiera eventuella urvalseffekter i den mekaniska urvalsmetoden och även potentiell drift. Det senare problemet med drift är inte i fokus för denna studie och inga vägningstal förekommer på varunivå utan enbart på produktgruppsnivå.

## 4 Resultat

Index har beräknats genom prisrelativer per produkt (aggregerat över ingående veckor per månad) och i samtliga butiker som ingick i KPI:s kassaregisterurval åren 2015 och 2016. Med begreppet *produkt* avses för fastkorgsindex den enligt KPI definierade produktkoden och är inte påverkad av att EAN-koder byts ut under året – produkten följs medan innehållet kan ändras. För månadsmatchningsmetoden görs beräkningen genom matchning av varje EAN-kod. Detta gäller även för den kontrollerade metoden.

Det totala antalet produkter i månadsmatchningen uppgår till i genomsnitt c:a 1100 produkter (EAN-koder) per månad i COICOP 01 och överstiger därmed de motsvarande cirka 750 produkter som ingår i fast korg. Av dessa är knappt 600 produkter sådana som ingår i COICOP 01. Dock bör noteras att antalet i en specifik produktgrupp kan vara färre i den mekaniska ansatsen än i fastkorgsansatsen.

Diagram 1a och 1b visar index för fastkorgsindex och månadsmatchning med automatiskt urval, för respektive år. För båda åren ligger fastkorgsindex över månadsmatchningen. Med utgångspunkt från resultaten analyserades datamaterialet och det kunde identifieras att månadsmatchningen, i den form som används här, innehåller produkter vars priser påverkas kraftigt av utförsäljningar/kampanjer som är påtagligt i januari 2016.

**Diagram 1a** Fastkorsindex och månadsmatchning, COICOP 01, år 2015.**Diagram 1b** Fastkorsindex och månadsmatchning, COICOP 01, år 2016.

**Not:** Vertikala axelvärden är borttagna för att inte röja indexnivåerna.

En annan tydlig skillnad mellan metoderna kan ses under period 4, år 2016 då de två indexserierna visar olika riktning. Under denna period sker en förändring i varugrupsbeteckningarna och det är troligt att förändringen får en effekt på urvalet men detta stabiliseras efter den månad det inträffar eftersom urvalen

uppdateras automatiskt och matchas en månad bakåt.

En tidigare beräkning av medelfelet inom dagligvaror genom kassaregisterdata (ej publicerad) har visat på att osäkerhetsbidraget från produkturvalet betingat på det befintliga butiksurvalet blir 0,168 indexenheter i genomsnitt per månad. Ställt i relation till diagrammen kan ses att de två metoderna inte avviker tydligt förutom på grund av två enskilda avvikelser, en i april 2015 och en i januari 2016.

För att visa på eventuella skillnader mellan de två metoderna och kontrollera för eventuella urvalseffekter har följande två variabler härletts:

- a) Differens mellan kontrollerad månadsmatchning och okontrollerad månadsmatchning
- b) Differens mellan fastkorgsindex och kontrollerad månadsmatchning

De härledda variablerna, differenserna  $a$  och  $b$ , kan tolkas enligt följande.

**Differens a)** ger avvikelserna för automatisk månadsmatchning från kontrollerad månadsmatchning. Detta är den skillnad som kan tänkas uppstå av att ha en helt mekanisk ansats i den månatliga urvalsdragningen och att urvalet generellt blir större, istället för att använda ett helt kontrollerat och mindre urval och tillämpa månadsmatchning.

**Differens b)** ger avvikelserna för kontrollerad månadsmatchning från fastkorgsindex. Detta kan indikera på eventuella metodeffekter och inte urvalseffekter.

I Tabell 2 visas på differensernas utfall för åren 2015 och 2016.

**Tabell 2** Medelvärde av månatliga differenser  $a$  och  $b$ , åren 2015-2016.

År	Differens (a)	Differens (b)
2015	0,105	0,359
2016	0,793	0,033

Förhållandet mellan differenserna  $a$ ) och  $b$ ) är olika de två åren enligt Tabell 2. För 2015 kan ses att avvikelserna i  $a$ ), urvalseffekten, är liten samtidigt som  $b$ ) är stor, vilket pekar på en metodeffekt av att jämföra parvisa månader. För 2016 ses motsatt effekt, nämligen att urvalseffekten i  $a$ ) är stor medan kontrollerad månadsmatchning inte avviker från fastkorgsindex.

## 5 Diskussion och fortsatta studier

Ett intressant resultat från studien är att KPI:s fastkorgsindex ligger över månadsmatchningen. Eftersom analysen avser en begränsad period, 24 månader, bör resultaten tolkas försiktigt, men två särskilda effekter noterades i materialet som gav upphov till skillnader mellan metoderna. Den första effekten var från att urvalet i månadsmatchningen för januari 2016 påverkades i stor utsträckning av

många låga priser. Detta betyder att någon form av prisdumpningsfilter kan behöva användas i månadsmatchningen.

Den andra effekten som noterades är ändrad kodning, vilket leder till en instabilitet i urvalet till månadsmatchningen om det inte beaktas direkt. Förändringar i kedjornas varugrupper innebär att antal produkter i en produktgrupp behöver bevakas eftersom klassificeringen av produkter kan ändras över tid. En möjlig effekt av detta i studien är att produkter kan ha hamnat i fel produktgrupp vilket innebär att produkten får ett felaktigt vägningstal i aggregeringen till COICOP 01.

Allt fler länder använder kassaregisterdata i sina KPI/HIKP-beräkningar. Med tanke på att samtliga av dessa använder andra metoder jämfört med SCB ökar behovet av ytterligare analys för att säkerställa att vald metod är den bästa möjliga. Denna studie är ett första steg i den riktningen.

## 6 Referenser

Grient, van der H. & Haan, J. de. (2010). The use of Supermarket data in the Dutch CPI.

## 7 Appendix

### 7.1 Tröskelvärdesbestämning

För ett elementäragregat  $a$  finns det försäljning för  $N$  olika EAN-koder månaderna  $m$  och  $m-1$  i år  $y$ :  $N_a^{(y,m),(y,m-1)}$ . Om försäljningen för en viss EAN-kod  $i$  för månad  $m$  uttrycks som en andel  $s_i^{y,m}$  av produktgruppens totala försäljning och där genomsnittsandelen för de båda ingående månaderna överstiger ett visst tröskelvärde  $B$ ,  $\frac{s_i^{y,m} + s_i^{y,m-1}}{2} > B$  inkluderas EAN-koden i urvalet  $n_a^{(y,m),(y,m-1)}$ .

*Kommentar: imputeringar*

En förekommande hantering vid bortfall av prisobservationer är imputeringar för att upprätthålla stabilitet i ansatsen. För denna förenklade studie har ingen bortfallshantering skett då det skulle krävas en genomtänkt metodstruktur. I detta avseende avviker därför ansatsen här från andra länders implementeringar genom att endast observerade data har inkluderats. Detta i sig torde inte påverka jämförelsen mot fastkorgsindex.

### 7.2 Aggregering över uppgiftslämnare

Den empiriska tillämpningen i denna studie, enligt formuleringarna ovan, görs per uppgiftslämnare till kassaregisterdata. Detta är analogt med den nederländska tillämpningen där ett index, elementäragregat, skapas per uppgiftslämnare (dagligvarukedja). För att kunna ta fram det slutliga indexet per produktgrupp  $a$  görs sammanvägning av de ingående uppgiftslämnarna enligt deras omsättningsandelar. De ingående indexen vägs ihop geometriskt till produktgruppsindex.