

Vägningstal för färsk frukt och färska grönsaker från kassaregisterdata

För diskussion

Från och med år 2017 har användningen av kassaregisterdata utökats till att även täcka färsk frukt och färska grönsaker från en av de tre uppgiftslämnarna för dagligvaror. En principiell fråga som behöver diskuteras är hur varianter inom en produkt, exempelvis olika tomater ska vägas samman i prismätningen när det samtidigt görs manuell prisinsamling i övriga butiker. Denna PM syftar till en principdiskussion kring hur sammanvägningar av produktbjudanden inom en produktgrupp kan göras i fallet med variation i utbudet och kända konsumtionskvantiteter.

INNEHÅLL

1	BAKGRUND	1
1.1	DEFINITION AV PRODUKT, PRODUKTGRUPP ETC.	2
1.2	MANUELL INSAMLING I BUTIK	2
1.3	KASSAREGISTERDATA	2
2	PROBLEM OCH MÖJLIGHET	3
3	OLIKA TÄNKBARA URVAL OCH SAMMANVÄGNINGAR	3
3.1	BEFINTLIGT URVAL OCH VÄGNINGSFÖRFARANDE	3
3.2	MÖJLIGA URVAL OCH VÄGNINGSFÖRFARANDEN.....	3
3.3	REKOMMENDATIONER ENLIGT HIKP	5
3.4	UTBLICK: HUR GÖR ANDRA LÄNDER?	5
4	DISKUSSION	6
5	REFERENSER	6

1 Bakgrund

På nämndens sammanträde nr 1, oktober 2016 lyfte SCB frågan om införande av kassaregisterdata för färsk frukt och färska grönsaker från och med 2017. Nämnden tillrådde hantering enligt existerande metoder med nödvändiga praktiska avvägningar.

Mot denna bakgrund har priser för färsk frukt och färska grönsaker börjat mätas genom kassaregisterdata för en av de tre dagligvarukedjorna som levererar data till KPI. Därmed har manuell insamling avslutats för den dagligvarukedjan men bibehållits för övriga två



dagligvarukedjor. Inom färsk frukt och färska grönsaker används så kallade PLU-koder (Product Look-Up) för att identifiera produkter, utöver vissa förpackade sådana varor där det finns EAN-koder. Eftersom förutsättningen för att införa kassaregisterdata är och har varit att det ska fungera i symbios med manuell prisinsamling har det krävts vissa avvägningar i samband med införandet av denna datakälla.

1.1 Definition av produkt, produktgrupp etc.

KPI har definierat totalt c:a 360 *produktgrupper*, varav 26 inom färsk frukt och färska grönsaker. En produktgrupp skall ha ett vägningstal, baserat på försäljningen år t-2. Vägningstalen är ofta tämligen skönsmässigt satta för frukt och grönsaker. Vissa frukter och grönsaker har delats in i två eller flera produktgrupper. Potatis är ett exempel på detta, med produktgrupperna Potatis, paketerad i påse (0,2 ‰), Potatis lösvikt (0,9 ‰) och Potatis, kravmärkt (0,1 ‰). Tomater finns däremot inom en enda produktgrupp (1,9 ‰).

En produktgrupp omfattar en eller flera *produkter*. Produkter har en av SCB fastställd definition som endera är generisk (begränsat storleksomfång, prestanda, ursprungsland etc.) eller specifik (exakt artikel avses, identifierbar med t.ex. EAN-kod). Produkter kan ha olika vägningstal. Produktgruppen tomater har till och med 2016 haft 7 produkter definierade:

- Tomat A, lösvikt med kvist
- Tomat B, lösvikt utan kvist
- Tomat C, Körsbärs-/cocktailtomater, 250 g
- Tomat D, Kravmärkta med kvist
- Tomat 5, sannolikhetsurval av förpackad tomat med EAN-kod i kassaregisterdata
- Tomat 6, sannolikhetsurval av förpackad tomat med EAN-kod i kassaregisterdata
- Tomat 7, sannolikhetsurval av förpackad tomat med EAN-kod i kassaregisterdata

I ca 40 butiker med lokal prisinsamling skall intervjuaren välja ett *produkterbjudande*, en variant, av vardera av produkterna A, B, C och D. I ca 40 andra butiker används kassaregisterdata med sannolikhetsurval av varianter, med urvalssannolikhet proportionell mot försäljningsvärde år t-2. Vi tänker oss här att vägningstal för de tre valda varianterna (per butikskedja) beräknas som produkten av variantens försäljningsvärde och den inverterade urvalssannolikheten, vilket blir 1.

Vid beräkning av prisindex för elementärt aggregat skapas ett ovägt geometriskt medelvärde av alla observerade priskvoter. I en butik med manuell datainsamling kan finnas en eller flera varianter (A, B, C och D), i butiker med insamling från kassaregisterdata kan vi istället hitta varianterna 5 - 7. Alla observerade produkterbjudanden ges alltså explicit samma vikt.

1.2 Manuell insamling i butik

Manuell insamling av priser görs per enskild produkt som är definierad enligt en beskrivning t.ex. "Tomat A, med kvist". Prisinsamlaren väljer själv en viss variant utifrån det som finns tillgängligt i butiken för säsongen. Mest varje månad kan prisinsamlaren behöva avgöra vad som verkar vara mest såld variant, d.v.s. representativ. Vanligtvis mäter han/hon ändå samma variant i efterföljande perioder om det går.

1.3 Kassaregisterdata

För kassaregisterdata avseende förpackade produkter, som produkterna Tomat 5 – 7, ser vi ett problem med att försäljningsuppgifter används i implicita vägningstal till skillnad från ovägda produkter med manuell insamling. Problemet är de blandade förfarandena.

Med kassaregisterdata för lösviktsvaror av frukt och grönt har vi nya möjligheter – och problem – med hur de skall väljas och eventuellt viktas vid aggregering till produktgrupp.

2 Problem och möjlighet

PLU-koden är den identifikation vi kan använda på ett rationellt sätt för kassaregisterdata för färska frukter och grönsaker. En förutsättning är nu att förekomsten av olika PLU-koder inom produktgrupper är så varierande från tid till annan att det inte fungerar att välja varianter år t-2 för att prismätas respektive månad under år t. Det är till och med så att det finns olika varianter mellan butiker inom veckor och mellan veckor för butiker. Dessutom påverkar säsongförhållanden utbudet. Detta kräver en sammanvägning av prisobservationerna på ett sätt som inte förekommer för övriga produkter inom ramen för den befintliga konstruktionen av elementäraggregat.

Den rika datamängden i kassaregisterdata tillåter att stora urval av produkter/produktbjudanden kan användas, rent av totalundersökningar. Nu uppstår frågan om hur ett stort urval i kassaregisterdata ska aggregeras och hur dessa data ska kombineras med befintliga *unika* mätningar i manuell prisinsamling.

3 Olika tänkbara urval och sammanvägningar

3.1 Befintligt urval och vägningsförfarande

I ett inbyggt experiment används kassaregisterdata under 2017 för färsk frukt och färska grönsaker i en butikskedja. Ett urval för butikskedjan har gjorts genom att identifiera de två eller tre mest sålda *produktvarianterna* per produkt, exempelvis tomater utan kvist, underföregående år.

Sammanvägning av produktvarianter och veckor är en summering av omsättningsbeloppen för varianterna den aktuella månaden, delat med sammanlagd kvantitet för månaden, vilket ger ett pris per butik och produkt. Vägningsstal används inte från föregående år eftersom effekten av säsongvariation inte är utredd. Vidare bör kanske inte kvantiteter eller värden för december år t-1, d.v.s. basperioden, användas i eventuella vägningsstal.

3.2 Möjliga urval och vägningsförfaranden

Tre huvudsakliga metoder för urval kan vi identifiera inom färsk frukt och färska grönsaker. För sammanvägningen av varianterna av produkter ser vi också ett antal möjligheter, se fallen 1-3, med respektive varianter a-c i tablå A.

I samtliga fall ligger ett vägningsproblem i aggregeringen från enskilda noteringar per vecka till en månadsprisnotering. Den information som finns i kassaregisterdata är, per variant i en viss butik, *total omsättning* och *försåld kvantitet*, det senare vanligen i kilogram, för denna typ av varor men även antal styck för varor som inte vägs utan säljs enhetsvis. Med ”mest såld” menar vi i följande tablå den som har störst försäljningsvärde i kronor.

Tablå A Tänkbara urvalsmetoder för varianter inom en produkt samt metoder för beräkning av pris per butik, produkt och månad i kassaregisterdata.

Fall	Urvalsmetod för varianter inom en produkt	Beräkning av pris per butik, produkt och månad
1a	Välj de n mest sålda varianterna per butik under mätveckan	Aritmetiskt ovägt medelvärde av kilopriser för produktvarianter och veckor
1b		Aritmetisk summa av försäljningsvärde dividerat med summa kvantiteter, d.v.s. kvantitetsvägt aritmetiskt medelvärde av kilopriser för produktvarianter och veckor
1c		Geometriskt ovägt medelvärde av kilopriser för produktvarianter och veckor
2a	Välj de n mest sålda varianterna per butik under månaden (summera veckor först)	Aritmetiskt ovägt medelvärde av kilopriser för produktvarianter
2b		Aritmetisk summa av försäljningsvärde dividerat med summa kvantiteter, d.v.s. kvantitetsvägt aritmetiskt medelvärde av kilopriser för produktvarianter
2c		Geometriskt ovägt medelvärde av kilopriser för produktvarianter
3a	Välj de k mest sålda varianterna enligt ett förutbestämt gemensamt urval för hela butikskedjan	Aritmetiskt ovägt medelvärde av kilopriser för produktvarianter och veckor
3b		Aritmetiskt vägt medelvärde av kilopriser för produktvarianter och veckor, med vikter från år $t-1$
3c		Aritmetisk summa av försäljningsvärde dividerat med summa kvantiteter, d.v.s. kvantitetsvägt aritmetiskt medelvärde av kilopriser för produktvarianter och veckor

Ett förutbestämt urval dras förslagsvis baserat på försäljning år $t-1$, möjligen exklusive december månads data som ju ingår i basen. Fall 3c motsvarar den metod för sammanvägning av varianter från kassaregisterdata som har införts från 2017. Om årsvikter används (fall 3b) torde dessa behöva avspegla närmast föregående år för att säkerställa att befintligt urval faktiskt existerar i det kommande året. Om helårsvikter används och sista månaden, december, tillåts ingå innebär det att en del av vägningstalet motsvarar basperioden vilket avviker från befintlig indexkonstruktion. Därutöver uppstår ett problem eftersom när varianter upphör under året och behöver ersättas saknas årsvikter för den nya varianten.

I följande exempel illustreras fallen med medelpris viktat genom aktuell konsumtion visavi medelpris genom förutbestämda vikter när priset är konstant mellan två perioder men när konsumtionen förändras enligt preferenser.

3.2.1.1 Exempel: tomater utan/med förutbestämda vikter, ingen prisförändring

Antag att det finns två varianter av tomater *utan kvist*: tomat X och tomat Y. Tomat Y kostar 99 kr/kg och tomat X kostar 20 kr/kg. Eftersom motsvarande manuell insamling avser ett pris för *ett kilo tomater utan kvist*, som är den definierade produkten i det här fallet, måste de två varianterna X och Y i kassaregisterdata kombineras till *ett* pris avspeglade konsumtionen för dem båda.

I period 1 konsumeras 100 kg av tomat Y och 2 000 kg av tomat X.
I period 2 konsumeras 200 kg av tomat Y och 1 000 kg av tomat X.
I båda perioderna är priserna oförändrade, 99 kr respektive 20 kr.

Kilopriset för kombinationen i de två perioderna blir, vid sammanvägning enligt fallen 2 och 3 med veckovisa faktiska konsumtionsbelopp, följande vid oförändrat pris:

$$P(C,1) = (99*100 + 20*2000) / (100+2000) = 23,76$$

$$P(C,2) = (99*200 + 20*1000) / (200+1000) = 33,16$$

Med förutbestämda årsvikter i fall 8, t.ex. vikten 5 för B mot vikten 20 för A, blir kilopriset för kombinationen C konstant för de två perioderna vid oförändrat pris:

$$P(C,1) = (99*5 + 20*20) / (5+20) = 35$$

$$P(C,2) = (99*5 + 20*20) / (5+20) = 35$$

3.3 Rekommendationer enligt HIKP

I HIKP-regleringen (2013) för säsongprodukter så som frukt och grönsaker nämns följande:

- Variable weights may be attributed only at product level but the total weight per product group must be fixed and represent the average consumption over the year.
- Where seasonal products are not available for a certain period of time (out-of-season months), it is possible to change their weights to 0 for this period.
- In product groups containing seasonal products the weights for products that are in-season may vary only to the extent that it is necessary to allow for month-on-month changes in the composition of the basket.
- If all products were available all over the year, fixed weights would be used based on the average expenditures ...//... and the monthly patterns in product expenditures and in total expenditures would be disregarded.

3.4 Utblick: hur gör andra länder?

Enligt de skrivna källor vi kunnat ta del av via Internet förefaller det som att SSB (Norge) använder en MCR-metod (Monthly Chaining and Resampling) för hela kassaregisterdata-systemet men med fasta produktgruppsvikter, d.v.s. enligt HIKP-regleringen. Eftersom de uteslutande använder kassaregisterdata och månadskedjning för dagligvaror existerar inte vårt problem på samma sätt, se Rodriguez & Haraldsen (2005) eller Nygaard (2010).

Nederländerna använder *monthly chained matched-item index* för kassaregisterdata. Deras tillvägagångssätt innebär att korgen är dynamisk och således följer marknaden. En produkt inkluderas i den månatliga indexberäkningen om de genomsnittliga utgifterna överstiger ett givet värde. Ingen viktning görs på elementär nivå utan ett oviktat geometriskt medelvärde (Jevons) används, se Grient & Haan (2010).

I både Norge och Nederländerna används en principiellt annorlunda ansats med månadskedjning, till synes även inom säsongprodukter.

4 Diskussion

Denna PM syftar till en principdiskussion kring hur *produktvarianter* inom en *produkt* kan väljas och sammanvägas i fallet med variation i utbudet och kända konsumtionskvantiteter. På sikt kan vi hoppas att kassaregisterdata helt kan ersätta manuell insamling för livsmedel, men ännu ett tag måste manuell insamling göras för butiker för vilka SCB inte får kassaregisterdata.

1. Hur ska data från både manuell insamling och kassaregisterdata vägas samman?
2. Finns bästa förslag bland de nio fallen i tablå A?
3. För förpackade dagligvaror gör SCB sannolikhetsurval av cirka 800 produkter per butikskedja. Några andra länder använder i stort sett hela mängden kassaregisterdata. Bör SCB använda sig av urvalsundersökning för färsk frukt och färska grönsaker eller kan urvalsstorlekarna n respektive k sättas lika med populationsstorlekarna?

5 Referenser

Rodriguez, J. & Haraldsen, F. (2005). The use of scanner data in constructing elementary aggregates for food and beverages. Ideas and experiences from Statistics Norway.

Nygaard, R. (2010). Chain Drift in a Monthly Chained Superlative Price Index. Paper written for the joint UNECE/ILOs workshop on scanner data, Geneva, 10 May 2010.

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