

Kassaregisterdata i KPI – utökad tillämpning för dagligvaror

För beslut

Prisenheten föreslår att i KPI-beräkningarna från år 2013, ersätta intervjuarinsamlade priser med kassaregisterdata för dagligvaror exklusive färskvaror hos tre butikskedjor. Detta sker redan hos en av kedjorna under 2012, i enlighet med tidigare nämndbeslut. Förslaget innebär att användningen av kassaregisterdata i KPI-beräkningarna kommer att omfatta större delen av dagligvaruhandeln.

Bakgrund

I dokumentationen för KPI i systemet SCBDOK (offentligt tillgängligt på www.scb.se) beskrivs dagligvaruområdet enligt följande:

Med dagligvarusortimentet avses livsmedel, drycker, tobak, varor för hushållsunderhåll, personliga hygienartiklar m.m., dock ej färskvaror inom frukt och grönsaker. Färskt kött har sedan länge ingått i begreppet. För framtiden borde begreppet snarare omfatta motsvarande produkter med standardiserad förpackning som har EAN-kod, eftersom prismätningarna planeras att successivt hämta priser från s.k. butikskasseinformation.

Urvalen av varuvarianter har dragits ... med metoder för sannolikhetsurval. Detta möjliggörs av att de stora butikskedjorna vänligen bistår SCB med register över varuvarianter, med data om senaste års försäljningsvärde.

Förutsättningarna för att utnyttja kassaregisterdata (även kallade skannerdata) i prisinsamlingen på dagligvaror i KPI har tidigare behandlats av Nämnden för KPI. Detta skedde vid sammanträde nr 243-246 nämligen maj 2011, oktober 2011, april 2012 och oktober 2012. På mötet i maj 2011 redovisades en studie av Norberg m.fl. [2], och nämnden uttryckte sitt stöd för fortsatt arbete. På mötet i oktober 2011 beslöt nämnden att i KPI-beräkningarna för 2012 ersätta intervjuarinsamlade priser med kassaregisterdata för dagligvaror exklusive färskvaror hos en butikskedja.

På mötet i april 2012 konstaterades vid nämndens diskussion om rabatter i KPI att nämndens hållning var att rabatter bör inkluderas (dvs. vara avdragna i priserna som KPI räknas på) i så stor utsträckning som möjligt. Undantag är ”bäst före”-rabatter. På mötet i oktober 2012 konstaterades vid nämndens diskussion att transaktionspriser bör användas för de flesta produkttyper när de finns tillgängliga och att avsteg bör motiveras i varje enskilt fall.

Nämnden överlät till Prisenheten att avgöra vilka typer av rabatter som eventuellt ska exkluderas av praktiska skäl; nämnden önskade dock redovisning av sådana fall.

På mötet i oktober 2012 beslöt nämnden vidare att det månatliga genomsnittspriset per produkt och butik från kassaregisterdata ska räknas som kvantitetsvägt aritmetiskt medelvärde.

Tillgången till kassaregisterdata för KPI-beräkningarna säkras genom avtal med de tre butikskedjorna som förslaget berör. Avtalen är frivilliga för butikskedjorna och gäller tills vidare med 6 eller 12 månaders uppsägningstid.

Nuvarande tillämpning och förslag till utökad tillämpning

I användningen av kassaregisterdata för en butikskedja under 2012, i enlighet med nämndens ovannämnda beslut, utnyttjas kassaregisterdata som omfattar ett sannolikhetsurval av butiker och ett sannolikhetsurval av produkter. Urvalen är av samma design och storlek som i motsvarande intervjuarbase-rad insamling. Genom att datainsamlingen även för kassaregisterdata begränsas till ett urval, är det möjligt inte bara att hålla datamängderna hanterbara utan även att effektivt granska uppgifterna och beakta förekommande justeringar i olika EAN-koder m.m.

Användningen av kassaregisterdata för KPI-beräkningarna föreslås från 2013 utökas till tre butikskedjor. Den månatliga prisinsamlingen avses då ske från motsvarande urval som det nämnda, för dessa butikskedjor. Intervjuarinsamlingen av priser på dagligvarusortimentet från dessa butiker avses därmed upphöra och ersättas med kassaregisterdata. Detta gäller naturligtvis även insamlingen i december 2012 av baspriser för 2013. För den butikskedja där kassaregisterdata redan används i KPI avses dessutom antalet butiker och antalet produkter i urvalet fördubblas till 2013.

Ett syfte med förslaget är naturligtvis att fördelarna med kassaregisterdata (jfr. tidigare pm till nämnden av Sammar [4]) kan tas till vara fullt ut för prismätningen på dagligvarusortimentet.

Analysen för att säkra datamaterialens kvalitet och användbarhet redovisas sammanfattningsvis nedan. Även vissa ytterligare analyser av data ska göras så att förslaget kan genomföras med fullt säkerställd kvalitet i KPI-beräkningarna för 2013.

Genomförande

Dataöverföring

SCB har upprättat en säker dataöverföring från tre olika butikskedjor genom ett s.k. FTP-konto. Problemen som har uppkommit i samband med skapandet av FTP-kontot har snabbt kunnat redas ut och åtgärdas. Butikskedjornas olika system levererar artikeldata (bl.a. EAN, Artikelnamn, Snittpris och Antal) på veckobasis; se variabelförteckning i bilaga 1. Variablerna mellan butikerna benämns olika men vi får det vi efterfrågar. Leveransen sker alltid på onsdagar och vid den tidpunkt då data inkommer till SCB aviseras detta med ett automatiskt e-postmeddelande. Om butikskedjan inte levererat på avtalad tid finns möjlighet för SCB att återkoppla till butikskedjan.

Idag skickar butikskedjorna data i textfiler. SCB har hela tiden haft intentionen att underlätta arbetet för uppgiftslämnare. Därför har de fått skicka data till oss i det format de själva önskar. Leveranserna har gått till mycket smidigt. Ett problem som kan tänkas är om någon av butikskedjorna skulle börja skicka data i ett annat dataformat (t.ex. ANSI, ASCII, UTF-8, -16, -32) än dittills. Vi har identifierat teckensättningen för de olika butikskedjorna och skulle kunna hantera en ändring.

Produktionssystem

Kassaregisterdatasystemet för de tre olika butikskedjorna bygger på att vi får en textfil varje vecka som innehåller föregående veckas försäljning i utvalda butiker. I filen finns inte produktkoder utan EAN-nummer. Kassaregisterdatasystemet hanterar alltså dessa EAN-nummer och kopplar ihop dem med våra produktkoder. Allt detta görs i SAS-skript genom Pi09 (Prisenhetens produktionssystem).

Kassaregisterdatasystemet har fem övergripande produktionssteg:

1. Starta produktionsmånaden
2. Kontrollera inkomna data
3. Selekttera data och granska
4. Aggregera priser över tre veckor
5. Skicka data till Pi09

Dessutom finns ett steg som heter Marknadsanalys och där analyseras försäljningen av olika produkter. Flödet ser ut enligt följande: Marknadsanalyskörningen genererar en lista i Excel på alla EAN-koder som använts för den aktuella månaden och redovisar i hur många butiker som har haft försäljning av den varan. För att generera ett aggregerat pris för en butik räcker det med att en EAN-kod finns med i data för en av de tre mätveckorna. Priser jämförs sedan historiskt samt mellan butiker under aktuell vecka. Excel-listan går att sortera för att få fram de EAN-koder vars förekomst är

lägst jämfört med basmånaden december. Prisenheten har som regel att EAN-koder vars förekomst är 60 % eller lägre ska i marknadsanalysen undersökas vidare.

Det ska också noteras att marknadsanalysen görs mellan två perioder. Detta innebär att ändringar på EAN-koder en månad slår först under efterföljande månad.

I granskningsgränssnittet kommer priser ut som överskrider granskningsgränserna. När priser saknas historiskt sätts gränserna till 1 (Övre gräns=1, Undre gräns=1) och produkten måste granskas. För övriga produkter baseras gränser på innevarande periods uppgifter och även historiska uppgifter från kassaregisterdata. När det finns gränser kan granskningen göras selektivt.

Efter varje SAS-skript som körs i Prisenhetens gränssnitt görs en felsökning i de loggar som skapats, och även på de PDF-filer som genereras.

Sekretesskydd

Butiksdata är av mycket känslig karaktär och ägs ursprungligen av varje handlare. Den betingar ett stort kommersiellt värde, och därför har vi varit tydliga med att dataanvändandet kommer att begränsas till den offentliga statistikproduktionen. Butikskedjorna har informerats om SCB:s verksamhet och att data hos SCB skyddas genom reglerna i offentlighets- och sekretesslagen (2009:400) (OSL). SCB har även försäkrat butikskedjorna att uppgifterna som lämnas in till SCB för framställning av statistik även omfattas av sekretess enligt 24 kap. 8 § OSL (statistiksekretess). Sekretess gäller även enligt 31 kap. 16 § OSL för uppgift om en enskilds affärs- eller driftförhållanden när denne har trätt i affärsförbindelse med en myndighet, om det av särskild anledning kan antas att den enskilde lider skada om uppgiften röjs (affärssekretess).

Avtal

I och med att SCB får data utan kostnad har avtalet med butikskedjorna skrivits som en form av ett samarbetsavtal och inte ett avtal på kommersiella termer. Butikskedjorna frångår t.ex. sitt ansvar att ställas till försvar för kvalitén av data. Dessutom har SCB gått med på att servicen från butikskedjorna ska vara så begränsad som möjligt för att minska uppgiftslämnarbördan. I avtalet framgår dessutom att eventuella tvister i anledning av avtal skall slutligt avgöras genom skiljedom i enlighet med Stockholms Handelskammares Skiljedomsinstituts regler för förenklat skiljeförfarande. Kontaktpersoner knutna till avtalet och till data finns också att hitta.

Övrigt

I och med att dagligvarusortimentet inte innefattar viktvaror såsom frukt, grönt och färskt kött så avses heller inte kassaregisterdata för dessa produkter nyttjas i KPI. Ett nyttjande av kassaregisterdata för dessa produkter skulle förutsätta tillgång till kassaregistrerade kilopriser, något som inte verkar möjligt inom överskådlig tid.

Under arbetets gång har det dykt upp en del frågetecken kring data men med hjälp av butikskedjorna har alla kunnat redas ut.

Dataanalys på tillgängligt material

I den nämnda studien av Norberg m.fl. [2] redovisas avstämningar mellan medelpriser från kassaregisterdata med motsvarande priser insamlade manuellt av intervjuarna i butikerna. Den studien avsåg data från den butikskedja för vilken kassaregisterdata används i KPI under 2012.

Motsvarande studier görs nu för de två butikskedjor med vilka nyttjandet av kassaregisterdata i KPI föreslår utökas från 2013. Materialet omfattar tillgängliga priser under 2012 på ca 200 produkter i de butiker i kedjorna som ingår i KPI-urvalet för året.

Följande tabell ger en sammanställning av resultat från den tidigare studien (på en butikskedja här kallad "A") och från den nu genomförda studien på data för en av de enligt förslaget tillkommande butikskedjorna ("B").

Tabell 1. Procentuella fördelningar av utfall av jämförelser mellan medelpriser enligt kassaregisterdata (SD) och intervjuarinsamlade priser (MCP)

Rad nr	Utfall av jämförelsen	2009 jan-dec Butiker "A" Procent	2010 jan-dec Butiker "A" Procent	2012 apr-sep Butiker "B" Procent
1	Samma pris i SD och MCP	83,4	86,2	59,2
2	Lägre pris i SD än i MCP	4,3	3,7	17,6
3	Högre pris i SD än i MCP	4,8	3,3	4,8
4	Pris saknas i MCP, finns i SD	1,5	0,9	0,9
5	Pris saknas i SD, finns i MCP	4,5	5,3	14,1
6	Pris saknas i både SD och MCP	1,5	0,6	3,4
7	Totalt	100,0	100,0	100,0

Beteckningarna SD och MCP står för skannerdata resp. Manually Collected Prices och följer studien av Norberg m.fl. [2].

Priserna enligt kassaregisterdata (SD) är beräknade som genomsnittliga transaktionspriser under en vecka. De intervjuarinsamlade priserna är avlästa av intervjuarna på hyllkant eller motsvarande i butikerna vid mättillfället för KPI.

Kommentarer till utfallet. Den relativt stora andelen produkterbjudanden med lägre pris i SD än i MCP för butikerna "B" (rad 2 i tabellen) har en naturlig förklaring. Transaktionspriserna som SD avser är efter avdrag av flera slags rabatter än hyllkantspriserna i MCP. En teknisk skillnad mellan materialen från de två butikskedjorna är att SD för butikerna "A" även innehåller priser före avdrag av dessa rabatter.

Vidare kan noteras att priser relativt ofta kan saknas i SD, oftare än i MCP (rad 4-5, butikerna "B"). Även detta är naturligt, eftersom en vara som inte köps någon gång under veckan i butiken inte genererar något transaktionspris. Att detta förekommer oftare i butikerna "B" än "A" har troligen sin förklaring i det relativt lilla antalet butiker "B", där det finns en mindre "jourbutik" med begränsat sortiment.

Större diskrepanser har också gått igenom manuellt och delvis kunnat förklaras med t.ex. definitionsmissig otydlighet i intervjuarinsamlingen huruvida kilopris eller styckpris ska mätas. En sådan skiljaktighet i valet av enhet kan väntas lämna indexberäkningen opåverkad genom att den slår med samma faktor på täljare och nämnare, men den kommer fram i jämförelser på prisnivå.

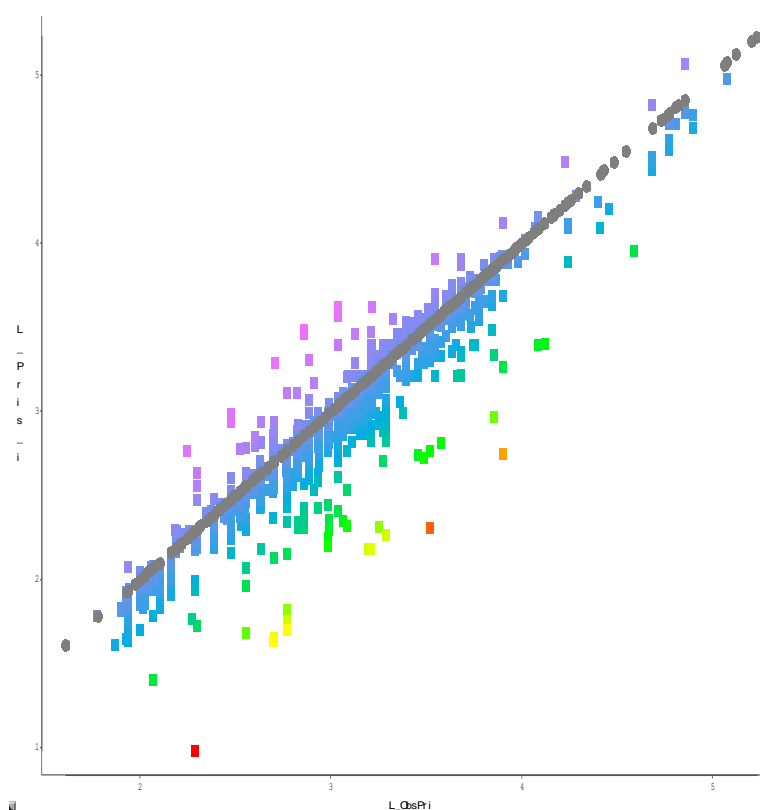
I ett relativt litet antal fall har kunnat noteras en mycket stor försäljningsvolym till ett pris som är lägre än hälften av medianpriset över tid för produkterbjudandet. En konsekvens av principen att beakta genomsnittligt transaktionspris är att sådana prisdykningar får relativt starkt genomslag i indexberäkningen. I produktion blir det därför önskvärt att verifiera dem, vilket kan väntas vara möjligt t.ex. genom att kolla butikernas reklamblad den veckan i fråga (finns tillgängliga på Internet). Troligen finns ofta en naturlig förklaring i kampanj- och lockpriser.

Vid jämförelser mellan SD och MCP ska hållas i minnet att en märkbar mätosäkerhet i intervjuarinsamlade priser har noterats i studier med upprepad intervjuarinsamling, såsom redovisas av Ribe [3]. Vidare har avvikelser i hyllkantspriser i butiker belagts i en rapport från Konsumentverket [1].

Sammantaget tyder inte resultaten på något alltför oväntat eller problematiskt i att gå över till att nyttja kassaregisterdata även i den föreslagna utökade

omfattningen. Enstaka ”problemvärden” som skulle kunna vara felaktiga och slå igenom i beräkningarna är så få att de kan väntas vara hanterliga i granskningen, särskilt i en produktionsmiljö i realtid med möjlighet till återkontakt med källan för verifiering.

I följande diagram är produkterbjudandena plottade efter intervjuarinsamlat pris på den horisontella axeln och pris enligt SD på den vertikala, för butikerna ”B”. Logaritmisk skala på båda axlarna. Åter ser man att fall med lägre kassaregistrerat pris än intervjuarinsamlat (dvs. punkterna under diagonal-linjen) tenderar att vara mera framträdande än de med motsatt förhållande.



Referenser

- [1] Konsumentverket, Prisinformation inom dagligvaruhandeln, Rapport 2010:2 (2009),
http://www.konsumentverket.se/Global/Konsumentverket.se/Best%C3%A4lla%20och%20ladda%20ner/rapporter/2010/2010_02_Prisinformation%20inom%20dagligvaruhandeln.pdf
- [2] A. Norberg, M. Sammar & C. Tongur, A study on scanner data in the Swedish consumer price index, Ottawa Group Meeting in Wellington, 2011,
<http://www.stats.govt.nz/ottawa-group-2011/agenda.aspx>
- [3] M. Ribe, Mode effect tests for handheld computers in price collection for Swedish CPI: Design and outcome, Draft, Statistics Sweden, 2009.
- [4] M. Sammar, Kassaregisterdata i KPI, Pm till Nämnden för KPI, 2011-10-11.

Bilaga 1

Variabellista

- Butiksnamn (Identifieringsnyckel)
- Postkod (Identifieringsnyckel)
- Period (År, vecka) (Identifieringsnyckel)
- EAN-nummer (Identifieringsnyckel)
- Artikelnamn
- Varumärke
- Förpackningsstorlek
- Antal sålda förpackningar
- Kvantitet med rabatt
- Omsättning per butik och EAN-kod, exklusive moms, där rabatter dras och depositionen är inkluderad
- Omsättning per butik och EAN-kod, inklusive moms, där rabatter dras och depositionen är inkluderad
- Omsättning per butik och EAN-kod, inklusive moms, där rabatter inte dras, men depositionen är inkluderad
- Pris inklusive moms
- Pris exklusive moms