

Hantering av extrema priser i kassaregisterdata

För information

I kassaregisterdata för dagligvaror fångas all försäljning per vara och butik. I det kan ingå kampanjer, reor och olika former av särskilda erbjudanden som gåvor eller varor med kort datum, där de två sistnämnda typerna av pris inte ska ingå i KPI/HIKP. Följden av detta är att priset i vissa fall kan bli påtagligt lågt. I denna PM presenteras beslutsregler för automatisk hantering av extrempriser.

INNEHÅLL

1	BAKGRUND	1
2	SYFTE	2
3	BEFINTLIG GRANSKNINGSMETOD FÖR MIKRODATA	2
4	FILTERANSATS FÖR EXTREMVÄRDEN	2
5	EXEMPEL PÅ ICKE-EXTREMA MEN AVVIKANDE PRISER	3
6	EFFEKTER AV DE OLIKA ANSATSERNA	5
7	FORTSATT STUDIER	5
8	REFERENSER	5

1 Bakgrund

En ökande datamängd från kassaregisterdata i KPI för dagligvaror har inneburit ökad manuell hantering och granskning. Exempelvis förekommer en del kraftigt avvikande (och tillfälliga) priser, huvudsakligen låga, som vanligen inte kan motiveras efter jämförande analyser med andra butiker, historiska data eller prisuppgifter på internet.

Kassaregisterdata som används i KPI innehåller veckovis omsättning i kronor och antal enheter, per EAN-kod och butik. Ibland blir det veckovisa enhetsvärdet påtagligt avvikande mot andra priser/historik och sådana priser påverkar månadsmedelpriset (aggregat över tre veckor) för produkterbjudandet i fråga.

Frågan om rabatter i kassaregisterdata har tidigare diskuterats på nämndmöte 246, november 2012, med anledning av underlaget ”Rabatter i kassaregisterdata för dagligvaror” (Sammar, 2012). Hanteringen av denna typ av datarelaterade frågor överlämnades till Prisenheten med förbehåll att redovisa för nämnden om ändringar görs (beslut enligt mötets protokoll).

Allteftersom Prisenheten har arbetat med kassaregisterdata har erfarits att de enhetspriser som ibland observeras möjligen kan reflektera olika situationer – kampanjpriser, rabatter men möjligen också gåvor eller icke-prima produkter, d.v.s. kort datum. Dessvärre finns inte



möjlighet att på löpande basis kontakta dataleverantörerna i realtid i nödvändig utsträckning då det inte ingår i träffade överenskommelser om kostnadsfri datafångst, samtidigt som de data som fås är aggregat från olika butiker inom samma kedja. Detta föranleder att vissa priser därför måste elimineras ur data i beräkningarna.

2 Syfte

Genom denna PM förklaras nuvarande metod för hantering av avvikande priser i kassaregisterdata (färsk frukt, grönsaker och kött/ost ingår ej i denna hantering) och möjlig övergång till att automatiskt identifiera och hantera dessa priser. En sådan standardiserad hantering kan anpassas genom objektiva kriterier men utelämnar en del andra avvikelser som exempelvis priser i kombination med försäljningskvantiteter som tillsammans inte verkar konsekventa, t.ex. icke-prima produkter (utförsäljning/kort datum).

3 Befintlig granskningsmetod för mikrodata

Granskningen av kassaregisterdata för dagligvaror görs på veckovisa observationer som en avvikelsekontroll av varje enskilt pris. Jämförelsepriset i granskningen är en sammanvägd median över historiska data och aktuella veckans data för den specifika produkten. Om enskilda priser överstiger medianen med en faktor två eller understiger halva medianen flaggas dessa som misstänkta och granskas. Som stödinformation finns kvantitetsuppgifter, alltså antal sålda enheter. En del kraftigt avvikande priser, som inte kan motiveras, elimineras. Detta görs bedömningsmässigt från fall till fall. Andra avvikande priser som inte bedöms lika extrema behålls vanligen.

Andelen prisnoteringar som flaggas enligt aktuell granskningsmetod varierar mellan 0,1-0,4 procent av alla observationer i kassaregisterdata för dagligvaror (drygt 100 000 per månad), vilket kan ses i tabell 1. De kraftigt avvikande priserna utgör i sin tur en varierande andel, mellan 10 och 40 procent av dessa. De som elimineras är misstänkta gåvor och icke-prima produkter.

4 Filteransats för extremvärden

I befintlig litteratur om hantering av kassaregisterdata inom dagligvaror återkommer *filter* för identifiering och hantering av avvikande priser. Exempelvis verkar nedan visade varianter av filter, med samma typ av toleransintervall, vara tillämpbara för att identifiera priser som direkt ska elimineras, d.v.s. extremvärden:

Filter 1 Prisförändringar månad-till-månad överstigande en faktor 4, alltså ökning 300% eller minskning med 75% mot månaden innan betraktas som felaktiga/orimliga och elimineras.

Filter 2 Prisförändringar relativt vägt medelpris från all prishistorik för produkten (EAN). Vägning baseras på försäljningskvantiteter vid varje pris. Enligt samma toleransintervall som första filtret minskning 75% (el. ökning 300%) mot medelvärdet betraktas som felaktiga/orimliga.

Filter 3 Prisförändringar relativt vägt medelpris från prishistorik för produkten i den specifika butiken. Vägning baseras på försäljningskvantiteter vid varje butik. Enligt samma toleransintervall som första filtret minskning 75% (el. ökning 300%) mot medelvärdet betraktas som felaktiga/orimliga.

Grundprincipen för filtret finns återgiven av van der Grient & de Haan (2010). Filteransatsen innebär egentligen en mekanisk formalisering av det som idag i viss utsträckning tillämpas manuellt/bedömningsmässigt för kassaregisterdata på Prisenheten. Utfallet på de tidigare eliminerade priserna visas i tabell 1.

Tabell 1. Total antal observationer från kassaregisterdata (2017) för dagligvaror, flaggat antal observationer och eliminerade observationer.

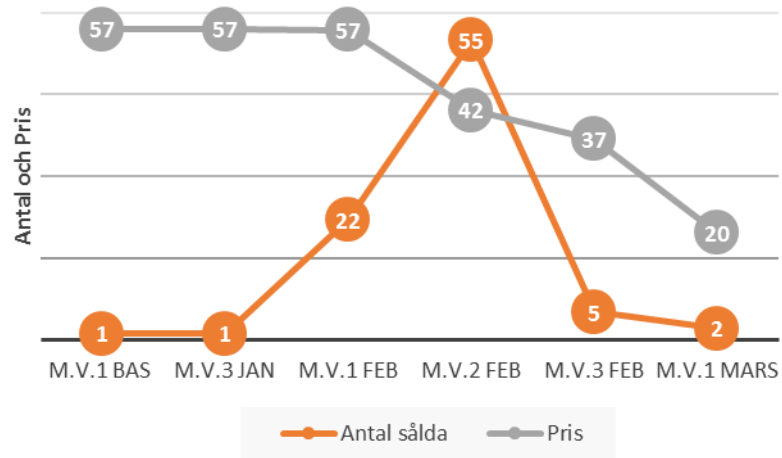
Period	Totalt	Flaggningar	Eliminerade (manuellt)	Eliminerade (filter)
1	111 188	406	49	16
2	109 199	240	73	45
3	107 765	164	43	16
4	107 434	135	37	25
5	106 520	127	35	13
6	106 645	158	39	12
7	105 668	119	37	12
8	106 157	126	25	14
9	105 474	174	36	13
10	104 542	222	39	25
11	103 778	202	28	27
12	104 095	192	68	32

Not: endast en handfull priser avviker uppåt i elimineringen.

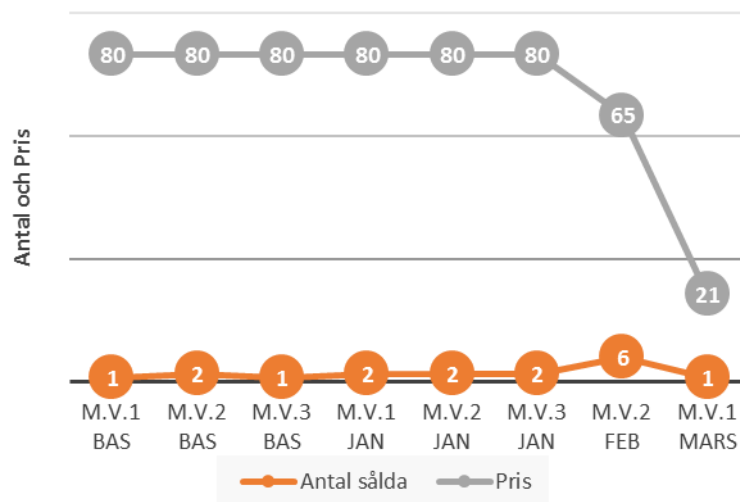
5 Exempel på icke-extrema men avvikande priser

Av de prisobservationer som har eliminerats i det empiriska underlaget finns en mängd som inte täcks in av filteransatsen. I vissa situationer är det svårt att bedöma huruvida en produkt har ett acceptabelt försäljningsbeteende eller inte. Generellt förväntar man sig att en produkts prishöjning följs av en minskning av den totala försäljningen, och tvärtom, en produkts prissänkning resulterar i en ökning av försäljningen.

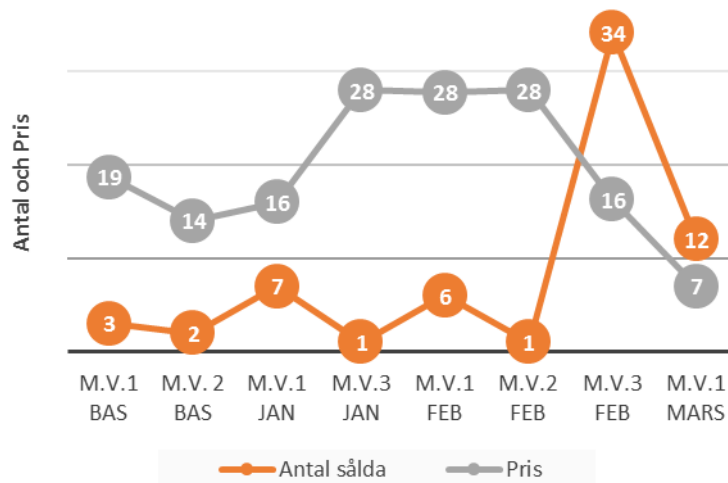
Följande (verkliga) exempel visar på förhållandet mellan en produkts pris och antal sålda enheter. Figur 1 till figur 3 återspeglar olika produkter i kassaregisterdata per mätvecka (1-3), butik och period.



Figur 1. Säsongsprodukt (allahjärtansdagschoklad)



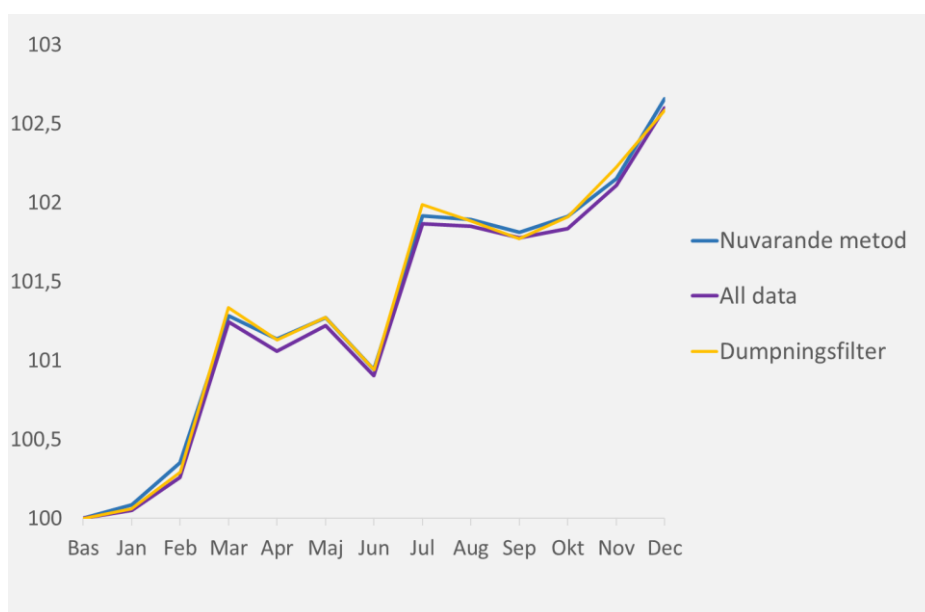
Figur 2. Produkt med konstant pris tom period 2 och konstant antal sålda (lammfärs)



Figur 3. Produkt som antas vara slutsåld period 3 (kanelbullar)

6 Effekter av de olika ansatserna

En tillämpning av den nuvarande granskningsmetoden och filteransatsen visas i figur 4 nedan där prisindex för 2017 beräknats för förpackade livsmedel (COICOP 01 exklusive färsk frukt, grönsaker, kött/ost). Samma beräkning utan någon som helst granskning visas också (lila linje). Som kan utläsas ur graf 4 är skillnaden mellan tillämpningarna liten på aggregerad nivå, dock kan underliggande produktgrupper visa större variation, i synnerhet sådana produktgrupper som har få prisobservationer.



Figur 4. Jämförelse av Index för dagligvaror på kassaregisterdata 2017-Nuvarande metod, dumpningsfilter och alla data (= utan eliminering).

7 Fortsatta studier

Prisenheten avser att fortsätta analysera effekten av beslutsregler för automatisk hantering av extrempriser. Vidare analys kommer göras på produktgruppsnivå.

8 Referenser

van der Grient, H., & de Haan, J. (2010). The use of supermarket scanner data in the Dutch CPI. CBS – Statistics Netherlands.

Sammar, M. (2012). Rabatter i kassaregisterdata för dagligvaror. PM till nämndmöte 246.