

## Hantering av urvalsbias i KPI

### *För information*

---

Prisenheten har antagit en stegvis ansats av implementering av den ”bästa” metoden för hantering av urvalsbias i KPI: tidigareläggande av prismätningar i nyvalda butiker till september. Inför 2015 prismäts branschen ”Radio- och TV-butiker” från och med september 2014. Under försöks-perioden kommer prisenheten att observera kostnader för metodändringen samt göra analyser av dess effekter. Dels kommer urvalsbiasen för de berörda butikerna att noteras, dels kommer variansförändringen till följd av mindre urval beräknas. Även om data-underlagen är små och osäkerheterna därmed stora anser vi att vi kommer ha tillräckligt med information för att fatta beslut om huruvida metoden kan anses vara lyckosam. Successiv implementering fortsätter i den mån som kvalitetsvinster bedöms stå i paritet med kostnader, allt i jämförelse med andra kvalitetsvinster som kan uppnås i KPI med motsvarande resurser.

---

## 1 Bakgrund

Under KPI-nämndens möte nr. 249 diskuterades den nedåtgående bias som uppstår då proportionen reavaror i urvalet är lägre i basmånaden för år  $y+1$  än i avslutande december månad år  $y$ . Detta problem uppmärksammades för första gången i produktgrupperna kläder och skor under 90-talet. Sedan dess har index för dessa produktgrupper korrigerats för att uppväga urvalsbiasen.

I den utredning som gjordes inför nämndens möte i april 2013 visade det sig att denna urvalsbias påverkar även andra produktgrupper. Flera produktgrupper, såsom högteknologiska produkter (ex. datorer), hemtextil (ex. dukar) och bosättningsaffärer (ex. glas och tallrikar) har realisationer under årets sista månader. Enheten presenterade inför nämnden olika möjligheter av hantering av urvalsbiasen.

## 2 Metod

### 2.1 Problem

Trots att instruktionen tydligt säger att alla typer av priser kan och skall inkluderas i prismätningarna, tenderar SCBs intervjuare att undvika produktvarianter med realisationspris



när de väljer produkterbjudanden i nyvalda butiker under basmånaden december. En bias uppstår i konsumentprisindex (KPI) på grund av intervjuarnas medvetna urvalsförfarande.

När urvalet är gjort och prisinsamling görs månad efter månad skall intervjuaren inte byta ut varor som realiseras så länge varorna fortfarande anses vara representativa. En vara är inte längre representativ om realisationen pågått flera månader i följd, om varianten flyttats till en plats med osorterade varor, endast udda exemplar finns kvar, eller om varan realiserats i flera steg. I december tolv månader efter introduktionen av det nya butiksurvalet kommer andelen sänkta priser därför normalt att vara högre än i basmånaden. Detta skapar en nedåtgående bias i KPI för berörda produktgrupper.

Ett parallellt problem uppstår med kampanjer, men i detta fall finns inga klara indikationer på att intervjuarna undviker eller favoriserar dessa vid sin val.

Det skall påpekas att endast de butiker som är nya i urvalet bidrar till den uppkomna biasen. Ett enkelt grepp skulle således vara att minska andelen nya butiker varje gång ett nytt urval dras. Det är emellertid SCBs policy att rotera 20 % av urvalet varje år för att minska uppgiftslämnarbördan. Dessutom försvinner mellan fem och tio procent av butikerna årligen på grund av förändringar i populationen, såsom konkurser och dylikt. Resultatet är att ungefär 30 % av butikerna byts ut årligen.

## 2.2 Förslag till lösning

Det finns två strategier för att hantera urvalsbiasen; försök att korrigera för den uppkomna biasen, alternativt; försök att hindra källan till bias.

### 2.2.1 Korrigera för bias

Korrigerig för uppkommen bias görs med kvoten av realisationernas (och/eller kampanjernas) effekter på observerade priser under en och samma period, i december då den trettionde prismätningen görs för innevarande år  $y$  och basen läggs för kommande år  $y+1$  med ett delvis nytt urval. Denna kvot är en skattning av bias som används för att korrigera index under det kommande året.

En tveksamhet till metoden kan motiveras av att "ordinarie priser" möjligen inte är tillförlitliga uppgifter i detaljhandeln. En analys av KPI-data 2010-2013 visar att varken prishöjningar eller prissänkningar är vanlig inför en realisation.

### 2.2.2 Förebygga bias

Man kan förebygga bias genom att dra nya butiksurval och låta intervjuarna göra urval av lokala produkterbjudanden i förslagsvis september, tre månader före den egentliga introduktionen av prisobservationerna i indexberäkningen. Från september till december kommer vissa av de intervjuarvalda varorna att realiseras och på så sätt minskar eller kanske helt elimineras biasen. Under 90-talet gjordes försök med denna metod då urvalet för vissa produktgrupper presenterades redan i november. Effekten var den önskvärda, men den var inte tillräckligt stark.

En nackdel med ett tidigare urval är att man måste prismäta två parallella urval för de aktuella produktgrupperna. Detta innebär givetvis mer jobb och högre kostnader för prismätningen.

Ytterligare en sätt att motverka uppkomsten av biasen är att för alla intervjuarinsamlade produktgrupper byta månad för urvalsuppdateringar till en månad med färre prisaktiviteter (såsom reor) än december. Om aktiviteter såsom kampanjer och realisationer är mindre fre-

kventa i t.ex. september än december blir kedjning av två indexlänkar med den överlappande månaden september av högre kvalitet, både i termer av bias och varians.

De reor som ändå finns under den nya länkmånaden kommer givetvis att ge samma bias som om urvalet länkas i december. Dessutom blir indexberäkningarna mer komplexa med denna metod.

### **2.2.3 Nämndens diskussion**

De tre förslagen till lösningar som presenterades inför nämnden var:

- 1) Dra urvalet för vissa produktgrupper i september för att tillåta proportionen av reor att öka innan urvalet används i KPI.
- 2) Byta länkmånad till en månad med färre prisaktiviteter, såsom reor och kampanjer.
- 3) Korrigera för den bias som uppstår med samma metod som idag används för kläder och skor.

Den sammanfattande bedömningen av nämndens diskussion är att biasen är så pass allvarlig att en åtgärd är motiverad. Det diskuteras huruvida en korrigering är alltför godtycklig bl.a. beroende på kvaliteten på de ordinarie priser som registreras. Huvudsakligen ifrågasätts om varan i fråga någonsin sålts till det ordinarie priset. Det finns farhågor att ordinarie priser används för att få reor att verka bättre än vad de egentligen är. Fördelen med en korrigering är å andra sidan att den kan införas omdelbart utan att belasta avdelningens ekonomi.

Att byta länkmånad till en månad med färre prisaktiviteter är inte intressant eftersom det inte tar hand om hela problemet.

Den mest attraktiva lösningen är att, för speciellt utsatta produktgrupper, dra ett urval i september som sedan länkas i december. Kostnaden med parallella urval skall givetvis ställas mot kostnaden med ett bias i KPI. Att flytta tillbaka urvalet för alla intervjuinsamlade produktgrupper anses både av nämnden och enheten vara en allt för dyr åtgärd. Ett bättre förslag är att börja med en produktgrupp som är särskilt påverkad av urvalsbias och sedan utvärdera hur stor effekten av ett tidigare urval blir.

### **2.2.4 Kläder och skor**

Ingen ändring i behandlingen av kläder och skor är planerad. Prisenheten kommer att fortsätta korrigera för att komma tillrätta med den relativt höga urvalsbias som uppstår varje år. Trots den kritik som framförts mot metoden anses den lösningen ändå vara den bästa. Biasen är så pass stor att en åtgärd anses nödvändig. Att dra urvalet tidigare för så stora produktgrupper skulle bli allt för tidskrävande och kostsamt.

## **2.3 Vidtagna åtgärder**

Inför 2015 har SCB börjat mäta priser i nya butiker inom branschen ”Radio- och TV-affärer” i september. Det är totalt nio butiker som omfattas av denna tidigare prismätning.

Urvalen för ”Bosättningsaffärer” och ”Hemtextilbutiker” har dragits ner något (totalt tre butiker färre) för att spara pengar och på så sätt förbereda för ett tidigare urval inför 2016 för dessa två branscher.

## 2.4 Biasskattningar och variansbidrag

Varje månad noteras 11 800 priser i det lokala prismätningssystemet. En tidigare mätning av Radio- och TV-affärer innebär ungefär 450 extra prisobservationer per månad (september – november). Eftersom sortimentskontrollen ej är nödvändig för dessa butiker kan man räkna med totalt ca 900 extra prisobservationer p.g.a. tidigareläggningen av prismätningen. Om dessa extra prisobservationer finansieras genom att urvalet minskar med motsvarande mängd betyder det en variansökning på 0,0000247, (variansskattningar från Nilsson, Ribe & Norberg, 2008).

Om vi utgår ifrån att MSE är ett bra mått på en undersöknings kvalitet vill vi avväga variansökningen mot kvadraten av den sammanvägda biasen för de aktuella produktgrupperna. Den totala vikten för de berörda produktgrupperna är 1,58 %. Ett geometriskt medel av biasen 2010 – 2014 för dessa produktgrupper är 1,5 %. Kvadraten av den sammanvägda biasen är 0,00056169. Slutsatsen är därmed att en lyckad åtgärd skulle sänka biasen mer än variansen förmodas öka.

Jämförelsen haltar emellertid något då variansskattningen är baserad på variansbidraget från alla lokalpriser. I lokalprisposten ingår såväl telefonundersökningar som besök. Eftersom Radio- och TV-butikerna enbart prismäts via besök och därför är något dyrare än den genomsnittliga lokala prismätningen måste urvalet förmodligen minskas något mer än med 900 observationer för att den ökade kostnaden skall kompenseras helt och hållet. Det bör också påpekas att beräkningarna är något osäkra då antalet observationer är lågt. Men, eftersom skillnaden mellan biasminskning och variansökning är så pass stor tycker vi ändå att utfallet är säkert.

## 3 Benchmarking

### 3.1 Meeting of the Group of Experts on Consumer Price Indices, UNECE

I maj 2014 presenterade Kristina Strandberg artikeln ”Correction of Sample Selection Bias” vid ”Meeting of the Group of Experts on Consumer Price Indices, UNECE” i Genève, Schweiz. Emottagandet var gott och ledde till intressanta diskussioner. Dels uttrycktes intresse från andra statistikbyråer av att kontrollera de egna urvalen för förekomsten av urvalsbias, dels kom en del förslag på hur man kan motverka uppkomsten av bias.

Även begreppen ”ordinarie priser” och ”reavaror” diskuterades. Fler länder än Sverige ser problem med dessa definitioner. Detta är i sig en stor fråga och behandlas inte vidare i denna text.

## 4 Slutsatser

Urvalsbias är ett problem som bör hanteras i KPI. Inför 2015 har prisenheten börjat prismäta produkter hos ”Radio- och TV-butiker” tidigare. Det rör sig om nio nya butiker där prisinsamlingen började redan i september 2014. Högteknologiska produkter (TV, CD, stereo, MP3-spelare osv.) har emellanåt hög urvalsbias och vi förväntar oss att det tidigare urvalet skall åtgärda detta.

Prisenheten har anammat en stegvis ansats och begränsar inför 2015 implementering av den ”bästa” metoden, tidigareläggning av prismätningar i nyvalda butiker till september, till en bransch. Prisenheten skall observera kostnader för metodändringen samt göra analyser av effekter, även om dataunderlagen är mycket små och osäkerheter därmed är stora. Successiv implementering fortsätter i den mån som kvalitetsvinster bedöms stå i paritet med kostnader, allt i jämförelse med andra kvalitetsvinster som kan uppnås i KPI med

motsvarande resurser.

## 5 Referenser

Nilsson, H., Ribe, M. och Norberg, A (2008). *Variansberäkningar KPI*. Projekt rapport, SCB, 2008-04-10.

Strandberg, Kristina & Norberg, Anders (2014). *Sample Selection Bias in the Swedish CPI*. Artikel presenterad vid Meeting of the Group of Experts on Consumer Price Indices, UNECE, 2014, 26-28 maj, Genève, Schweiz.