

Anders Norberg, Kristina Strandberg, Olivia Ståhl

# En modell för resursallokering inom KPI

*För diskussion*

---

KPI är en komplex statistikprodukt uppbyggd av ett 70-tal fristående undersökningar. Hur ekonomiska medel ska fördelas mellan dessa delundersökningar för att det totala måttet ska bli så bra som möjligt är därmed inte uppenbart. Dessutom har stora förändringar vad gäller indata skett under de senaste åren. Vi ser därför ett behov av ett hjälpmedel som kan ligga till grund för övergripande resursfördelningsval, dels mellan KPI:s olika delundersökningar och dels mellan olika åtgärder. Under 2017 har början till ett sådant hjälpmedel arbetats fram och denna PM utgör en beskrivning av såväl modellen som det pågående arbetet.

Nämnden för KPI välkomnas att diskutera arbetets inriktning och konsekvenser.

---

## Innehåll

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>BAKGRUND</b> .....   | <b>2</b>  |
| <b>2</b> | <b>SYFTE</b> .....  | <b>2</b>  |
| <b>3</b> | <b>TIDIGARE GENOMFÖRDA ARBETEN</b> .....  | <b>2</b>  |
| <b>4</b> | <b>EN MODELL FÖR FÖRDELNING AV RESURSER INOM KPI</b> .....  | <b>4</b>  |
| 4.1      | EN MODELL I TVÅ STEG .....  | 4         |
| 4.2      | STEG 1: VÄRDERING AV OSÄKERHETER OCH UPPSKATTNING AV KOSTNADER .....                                  | 5         |
| 4.3      | STEG 2: SAMMANVÄGNING OCH ALLOKERINGSFÖRSLAG .....  | 6         |
| <b>5</b> | <b>PRELIMINÄRA RESULTAT (PILOTSTUDIE)</b> .....   | <b>9</b>  |
| <b>6</b> | <b>DISKUSSION OCH FRÅGOR TILL NÄMNDEN</b> .....   | <b>9</b>  |
|          | <b>KÄLLOR</b> .....   | <b>10</b> |
|          | <b>BILAGA 1A: SKATTNING AV OMALLOKERINGSFAKTORER ENLIGT PROJEKTET "OPTIMAL ALLOKERING 2016"</b> ..... | <b>11</b> |
|          | <b>BILAGA 1B: OMALLOKERINGSFAKTORER ENLIGT PROJEKTET "OPTIMAL ALLOKERING 2016"</b> .                  | <b>12</b> |
|          | <b>BILAGA 2: MALL FÖR UTVÄRDERING AV BIASRISKER</b> .....   | <b>14</b> |



## 1 Bakgrund

KPI är en komplex statistikprodukt uppbyggd av ett 70-tal fristående undersökningar. Hur ekonomiska medel optimalt ska fördelas mellan olika delar av KPI är därför inte uppenbart. KPI har dessutom genomgått stora förändringar under de senaste åren, särskilt vad gäller indata, och vi ser inga tecken på att denna utveckling är på väg att stanna av. Kassa-registerdata, webskrapning och inköp av stora dataunderlag från externa företag är exempel på nya insamlingsmetoder som kommer att göra det nödvändigt att förändra hur vi handskas med data och vilka metoder vi använder för framställning av statistik.

Allokering av urval är en klassisk utmaning i all statistikproduktion. Oftast gäller allokeringsproblematiken emellertid en separat undersökning och den totala urvalskostnaden behandlas som exogent given. I KPI måste resurser allokeras mellan flera fristående delundersökningar. Dessutom kan andra osäkerhetskällor än urvalsfelet få stor betydelse för kvaliteten, vilket gör att det ibland är optimalt att fördela resurser från urval till andra kvalitetsdimensioner.

Sammantaget ser vi ett behov av ett hjälpmedel som kan vara till hjälp i den övergripande resursplaneringen, dels vad gäller fördelningen mellan KPI:s olika delundersökningar och dels mellan olika åtgärder. Under 2017 har början till ett sådant hjälpmedel arbetats fram, och denna PM utgör en beskrivning av det pågående arbetet.

## 2 Syfte

Denna PM beskriver det pågående arbetet med att utarbeta en modell som kan underlätta resursallokeringsbeslut inom KPI. Nämnden välkomnas att komma med synpunkter inför den fortsatta inriktningen.

## 3 Tidigare genomförda arbeten

Projekt med målsättningen att förbättra resursallokeringen inom KPI har bedrivits tidigare, men oftast med fokus på enbart urvalsosäkerheten. Skälet till att fokus legat på urvalsvarians är förstås att denna är mätbar och direkt relaterad till stickprovets storlek och därmed resurserna. Det är betydligt svårare att koppla olika typer av biasrisker till mängden tilldelade resurser då detta per definition kräver någon form av modellantagande.

En föregångare till föreliggande projekt genomfördes under 2016, även det med fokus på urvalsosäkerheten. Målet där var en modell för allokering av urvalsstorlekar mellan olika delundersökningar givet skattade urvalsosäkerheter, uppskattade rörliga kostnader förknippade med undersökningen, samt vägningstal. För att beräkna så kallade ”omallokeringsfaktorer” tillämpades en enkel modell för relationen mellan varians, kostnad och urvalsstorlek (se bilaga 1a för en beskrivning av modellen). Omallokeringsfaktorerna indikerar sedan vilka delundersökningar som bör tilldelas mer respektive mindre resurser. De faktorer som räknades fram under projektets gång redovisas i bilaga 1b.

Kartläggningar av biasriskerna i KPI har också gjorts vid flera tillfällen. SOU 1999:124, bilaga 6; *Bedömning av biasrisker i konsumentprisindex (KPI)*, har sin utgångspunkt i den så kallade "Boskin-rapporten" (Boskin et. al. 1996, 1998). I Boskin-rapporten diskuterades fyra olika typer av bias: *Substitutionsbias på hög nivå*, *Substitutionsbias på låg nivå*, *Nya produkter och kvalitetsförändringar* samt *Nya butiker*. I SOU-utredningen görs en liknande indelning men man tittar även specifikt på täckningsfel, mätfel och urvalsbias. Följande felkällor behandlas i utredningen (vi bortser för enkelhets skull från de delar som avser boendeposten specifikt):

- (1) Produktsubstitution, hög nivå
- (2) Produktsubstitution, intermediär nivå
- (3) Produktsubstitution, låg nivå
- (4) Nya butiker
- (5) Nya produkter
- (6) Kvalitetsförändringar
- (7) Täckningsfel
- (8) Mätfel
- (9) Urvalsbias

En uppdatering av denna studie gjordes även av Ribe (2017) inför KPI-nämndens höstmöte. 2012 gjorde prisenheten också en intern genomgång av risker för felaktigheter inom centralprisundersökningarna (se Andersson m.fl., 2012).

Inom svenska Producentprisindex (PPI) har man utarbetat en egen *indexvärderingsmodell*, vilken inkluderar såväl urvalsfel som icke-urvalsfel (se Draper och Schoultz, 2016; Draper och Fridén, 2017). Varje aggregat (SPIN) utvärderas var för sig och indexkvaliteten bedöms på en skala från 1-5. Bedömningarna sammanställs sedan till en beskrivning av kvaliteten inom varje SPIN. Modellen är tänkt att användas som underlag för förbättringsarbete och även som kvalitetsinformation till användare. Ingen sammanvägning av de olika aggregaten till en totalnivå görs dock, varför modellen som den ser ut idag inte kan användas för allokeringbeslut mellan olika delar av index.

Ett annat projekt som har viss koppling till detta är *Känslighetsanalys i BNP*, som genomfördes 2015-2016 (SCB, 2016a). En delaktivitet i det projektet var ett försök att bedöma risken för icke-urvalsfel i PPI och kombinera denna med variansen för att få ett mått på den totala osäkerheten i index. För varje biaskälla uppskattades risken för icke-urvalsfel på en skala mellan 1 och 3, och sedan kombinerades dessa skattningar med variansskattningarna utifrån ett antagande om icke-urvalsfelens totala storlek i förhållande till urvalsosäkerheten.

En viktig informationskälla är normalt sett även ILO-manualen (ILO, 2004). Vi har emellertid inte kunnat hitta så mycket information där som relaterar till projektet. Relevant är eventuellt ett stycke i kapitlet om kvalitetsjusteringar som handlar om vikten av *metadata*. Där står bland annat att metadatasystem kan vara till hjälp för att identifiera var kvalitetsvärderingsmetoder behöver ses över samt att bortfallsandelar bör följas över tid. Det

poängteras även att de delar som väger tyngst i KPI ska prioriteras för utveckling (en så kallad *top-down approach* bör användas).

Relevant för vårt arbete är slutligen även det kvalitetsbegrepp som från och med januari 2017 gäller för all officiell statistik i Sverige (SCB, 2016). Kvalitetsbegreppet har fem komponenter:

- Relevans
- Tillförlitlighet
- Aktualitet och punktlighet
- Tillgänglighet och tydlighet
- Jämförbarhet och sammanvändbarhet

I detta arbete har vi valt att fokusera på komponenten ”tillförlitlighet”, vilken rymmer följande osäkerhetskällor:

- Urval
- Ramtäckning
- Mätning
- Bortfall
- Bearbetning
- Modellantaganden

## 4 En modell för fördelning av resurser inom KPI

Med utgångspunkt i tidigare arbeten har en praktisk modell för resursallokering inom KPI utformats. Grundtanken har varit att hitta en metod att väga samman olika typer av osäkerheter, både systematiska och slumpmässiga, och att dessutom uppskatta deras effekt på KPI totalt. Skattade medelfel inom de olika delundersökningarna finns framtagna sedan tidigare. Dessa utgör främst skattningar av urvalsosäkerheten, men även andra typer av slumpmässiga fel kan påverka siffrorna. För att undvika risken för dubbelräkning och för att förenkla arbetet har vi valt att betrakta de skattade medelfelen som de enda slumpmässiga komponenterna i vår modell. För resterande osäkerhetskällor resonerar vi därmed i termer av bias. Detta antagande påverkar vårt sätt att jämföra de olika komponenterna med varandra eftersom vi väljer att relatera det kvadrerade medelfelet till summan av de övriga komponenterna.

I det preliminära arbetet har vi valt att använda KPI:s olika delundersökningar som indelningsgrund. Vilken detaljeringsgrad som är lagom är emellertid inte skrivet i sten och behöver sannolikt diskuteras närmare.

### 4.1 En modell i två steg

Den resursallokeringsmodell som tagits fram består av två steg. I det första steget värderas osäkerhetsriskerna i de delundersökningar som är föremål för utvärdering. Detta görs enligt en femgradig skala, där riskbetyg (RB) 1 innebär att biasrisken är så låg att inga åtgärder

behöver vidtas, medan riskbetyg 5 betyder att osäkerhetskällan medför en potentiellt stor bias. Kostnadsuppskattningar för att utföra förbättringar är också en del av modellens första steg, och ska anges i samtliga de fall där högsta betyg (riskbetyg 1) inte uppnås. Ett implicit antagande är att resurstilldelningen för delundersökningar som *inte* värderats i detta första steg redan är balanserad och att dessa undersökningar alltså inte är i behov av en översyn.

I modellens andra steg vägs riskvärderingarna samman. Utifrån sammantagen biasrisk, kostnadsuppskattningar, medelfel samt vägningstal skapas ett förslag på vilken åtgärd som bör prioriteras. Det kan exempelvis visa sig att urvalsstorleken bör utökas i en utpekad del av KPI, att externa data ska köpas in eller att ett specificerat utvecklingsprojekt bör prioriteras. Allokeringen kan i princip också visa att resurser för en viss del av KPI bör minskas, t.ex. genom att mängden granskning dras ner eller att en viss urvalsstorlek minskas (något som i så fall skulle visa sig genom att urvalsfelet för den delundersökningen hamnar långt ner i prioriteringslistan.)

Allokeringsförslagets kvalitet är förstått beroende av kvaliteten på input. För att modellen ska bli relevant krävs väl genomtänkta uppskattningar av osäkerhetsrisker och kostnader. Vår förhoppning är att modellen sedan ska göra det enklare att sammanfatta informationen och att den ska utgöra ett användbart underlag för den som ska ta beslut om slutlig resursfördelning.

## 4.2 Steg 1: Värdering av osäkerheter och uppskattning av kostnader

I det första steget vill vi få en uppskattning av storleken på olika typer av biasrisker i de enskilda delundersökningarna. Med inspiration från PPI:s indexvärderingsmodell har ett utkast till mall skapats för detta ändamål (se bilaga 2). Mallen utgör ett första utkast och förhoppningen är att den ska fyllas på och utvecklas allt eftersom, bland annat med mer fylliga beskrivningar och vägledningar för att förenkla för den person som ska fylla i den. Vi inser att utförliga vägledningar ökar sannolikheten för att få jämförbara uppskattningar för de olika delundersökningarna även om olika personer fyller i olika delar.

Till skillnad mot PPI:s modell väljer vi att använda de begrepp som ingår i SCB:s nya kvalitetsbegrepp (SCB, 2016b). Även om många av de kvalitetskomponenter och biasrisker som nämns i exempelvis SOU-bilagan och PPI-modellen är specifika för just prisindex går det bra att sortera in dem under rubrikerna i kvalitetsbegreppet. En annan skillnad mot PPI:s modell är att vi väljer att väga de olika kvalitetskomponenterna med olika vikter.

Enligt nuvarande förslag består mallen av 26 olika kvalitetskomponenter, urvalsfelet inkluderat. Det är tänkbart att vissa av dem skulle kunna slås ihop framöver för att få en mer lätthanterlig modell.

För varje delundersökning som ska utvärderas,  $u$ , och varje osäkerhetskomponent,  $o$ , bedöms risken för bias enligt skalan 1 till 5. Vi kallar detta värde  $RB_{u,o}$  ( $RB$  för *RiskBetyg*). Till varje osäkerhetskomponent förutom medelfelet hör även en implicit (i betydelsen att den inte syns i mallen) vikt, vilken vi valt att kalla  $Bet_o$  ( $Bet$  för *Betydelse*). Tanken är att man via parametern  $Bet_o$  ska kunna styra vilken inverkan en viss delkomponent har på den totala

biasrisken. Exempelvis kan man anta att det är av större betydelse för biasrisken om urvalet av produktbudanden är representativt än om butiksurvalet uppdateras ofta. Till att börja med tänker vi oss att  $Bet_o$  är fix mellan undersökningar, d.v.s. oberoende av  $u$ .

I samtliga fall där  $RB_{u,o} > 1$  ska även kostnaden för att minska risken uppskattas. Modellen är konstruerad så att den kostnad som ska uppskattas är kostnaden för att dra ner riskbetyget *ett steg*. För att kunna åstadkomma detta i praktiken har vi i pilotstudien använt oss av ett par förenklande regler. Den första regeln är att om en viss åtgärd kan ge upphov till en sänkning av riskbetyget med exempelvis två steg (säg, från 3 till 1) så har vi tagit hälften av den uppskattade totalkostnaden som vår kostnadsuppskattning. Den andra tumregeln vi använt är att vi betraktar rörliga kostnader under maximalt en femårsperiod. Om en viss åtgärd innebär en ökad månadskostnad så beaktar vi med andra ord endast  $5 \cdot 12$  sådana kostnadsökningar i vår totalkostnadsuppskattning.

För de osäkerhetskällor som har  $RB_{u,o} = 1$  finns även möjligheten att ange att *för mycket* resurser läggs på den aktuella delundersökningen. Kostnaden ska då anges med negativt tecken. Angiven kostnadsuppskattning tolkas i detta fall av modellen som den uppskattade besparing som kan göras om man godtar ett stegs försämring, dvs. går från risk 1 till 2.

Förutom osäkerhetsriskerna förknippade med olika typer av icke-urvalsfel ska en skattning av medelfelet anges i modellen. Även här måste en kostnadsuppskattning ges, men den kostnad som ska anges är en uppskattning av *de totala rörliga kostnader* som är förknippade med urvalet idag. Med andra ord, de kostnader som påverkas direkt av urvalsstorleken.

### 4.3 Steg 2: Sammanvägning och allokeringsförslag

När samtliga delundersökningar som ska vara föremål för utvärderingen värderats enligt mallen görs en sammanvägning för att få fram förslaget till omallokering. Här vägs kvalitetsbetyget och förbättringskostnaden (eller den möjliga besparingen) för de olika komponenterna samman och resultatet blir ett mått som indikerar vilken effekt man får av olika kvalitetsförbättrande aktiviteter (alternativt minskade resurser). Man kan sedan rangordna alla komponenter efter detta effektmått och på så sätt se vilka aktiviteter som ger mest förbättring för varje satsad krona.

Enligt samma princip som tillämpades i projektet Känslighetsanalys i BNP har vi valt att börja med att räkna om biasriskerna med en *omräkningsfaktor*,  $\omega$ , för att få dem jämförbara med medelfelet. I och med detta bestämmer vi att det totala medelfelet för KPI står i en viss relation till den totala biasen eller biasrisken. Notera att för vissa delundersökningar kommer medelfelet att vara noll då de beräknas via totalundersökningar. För andra undersökningar kan de uppskattade biasriskerna vara noll då de har fått betyget 1 för samtliga komponenter.

Låt  $U$  beteckna antalet delundersökningar i utvärderingen och  $O$  antalet kvalitetskomponenter. Vi sätter upp följande uttryck för sammanlagt medelfel och

sammanlagd biasrisk:

$$\text{Medelfel} = \sqrt{\sum_{u=1}^U \text{Vtal}_u^2 \cdot D_u^2}$$

$$\text{Biasrisk} = \omega \cdot \sum_{u=1}^U \text{Vtal}_u \sum_{o=1}^O \text{Bet}_o \cdot f(\text{RB}_{u,o})$$

där  $\text{Bet}_o$  är den relativa betydelsen av kvalitetskomponent  $o$  jämfört med övriga komponenter, vilken vi valt att konstruera så att  $\sum_{o=1}^O \text{Bet}_o = 1$ ,  $\text{Vtal}_u$  är vägningstalet för delundersökning  $u$ ,  $D_u$  är det skattade medelfelet förknippat med delundersökning  $u$  och  $f(\text{RB}_{u,o})$  är en monotont växande funktion av  $\text{RB}_{u,o}$ , t.ex.  $f(\text{RB}_{u,o}) = (\text{RB}_{u,o} - 1)$ ,  $f(\text{RB}_{u,o}) = (\text{RB}_{u,o} - 1)^Y$ , eller varför inte  $f(\text{RB}_{u,o}) = 1,1 + 1,25 \cdot \text{RB}_{u,o} - 0,2 \cdot \text{RB}_{u,o}^2 + 0,05 \cdot \text{RB}_{u,o}^3$ .<sup>1</sup> Vi kan nu bestämma värdet på  $\omega$  exempelvis på så vis att den sammanlagda biasrisken är lika stor som det sammanlagda medelfelet, vilket ger:

$$\omega = \frac{\sqrt{\sum_{u=1}^U \text{Vtal}_u^2 \cdot D_u^2}}{\sum_{u=1}^U \text{Vtal}_u \sum_{o=1}^O \text{Bet}_o \cdot f(\text{RB}_{u,o})}$$

Vi kan välja ett lägre värde på  $\omega$  om vi exempelvis har skäl att tro att de flesta osäkerhetskomponenterna ger upphov till slumpmässiga fel i större utsträckning än systematiska och därmed bidrar till det skattade medelfelet, eller att bias från de olika delundersökningarna "tar ut varandra".

För att komma fram till en modell för hur resursallokeringen bör se ut, på marginalen, resonerar vi så här: Antag att en krona ska tillföras någon av de aktuella delundersökningarna. Med den kronan skulle vi antingen kunna köpa oss ett lägre medelfel eller en lägre biasrisk. Antag först att vi väljer att lägga kronan på att utöka urvalsstorleken i undersökning  $a$ . Det totala medelfelet kommer då att ändras till:

$$\text{Medelfel}^+ = \sqrt{\sum_{u=1, u \neq a}^U \text{Vtal}_u^2 \cdot D_u^2 + \text{Vtal}_a^2 \cdot \frac{K_{a,c}}{(K_{a,c}+1)} \cdot D_a^2}$$

där  $K_{a,c}$  betecknar den kostnad som är förknippad med *medelfelet* i delundersökning  $u$  enligt modellens steg 1.

<sup>1</sup> I pilotstudien användes funktionen  $f(\text{RB}_{u,o}) = (\text{RB}_{u,o} - 1)$ .

Tack vare den extra kronan kommer medelfelet att förändras enligt:

$$[\text{Medelfel} - \text{Medelfel}^+] = \quad (1)$$

$$\sqrt{\sum_{u=1}^U \text{Vtal}_u^2 \cdot D_u^2} - \sqrt{\sum_{u=1, u \neq a}^U \text{Vtal}_u^2 \cdot D_u^2 + \text{Vtal}_a^2 \cdot \frac{K_{a,c}}{(K_{a,c}+1)} \cdot D_a^2}$$

Antag nu istället att vi väljer att lägga den extra kronan på att minska biasrisken  $b$  i delundersökning  $a$  med ett steg. Vi måste nu införa antagandet att biasrisken minskar proportionerligt med varje tillförd krona. Med andra ord antar vi att även om den totala kostnaden för att utföra en viss åtgärd är större än en krona, så kan vi minska biasrisken med motsvarande andel av kostnaden. Efter den tillförda kronan får vi då en ny biasrisk på formen:

$$\begin{aligned} \text{Biasrisk}^+ = & \sum_{u=1}^U \text{Vtal}_u \sum_{\substack{o=1, \\ o \neq b \text{ när } u=a}}^O \text{Bet}_o \cdot \omega \cdot f(\text{RB}_{u,o}) \\ & + \text{Vtal}_a \cdot \text{Bet}_b \cdot \omega \cdot \left[ \left( \frac{K_{a,b}-1}{K_{a,b}} \right) f(\text{RB}_{a,b}) + \left( \frac{1}{K_{a,b}} \right) f(\text{RB}_{a,b} - 1) \right] \end{aligned}$$

Skillnaden i biasrisk till följd av den extra kronan kan skrivas:

$$[\text{Biasrisk} - \text{Biasrisk}^+] = \text{Vtal}_a \cdot \text{Bet}_b \cdot \omega \cdot \left[ \frac{f(\text{RB}_{a,b}) - f(\text{RB}_{a,b}-1)}{K_{a,b}} \right] \quad (2)$$

Antag slutligen att vi istället skulle vilja utvärdera en möjlig besparing på marginalen. Om vi exempelvis väljer att spara en krona genom att öka biasrisk  $b$  inom delundersökning  $a$  så landar vi i den nya biasrisken:

$$\begin{aligned} \text{Biasrisk}^- = & \sum_{u=1}^U \text{Vtal}_u \sum_{\substack{o=1, \\ o \neq b \text{ när } u=a}}^O \text{Bet}_o \cdot \omega \cdot f(\text{RB}_{u,o}) \\ & + \text{Vtal}_a \cdot \text{Bet}_b \cdot \omega \cdot \left[ \left( \frac{K_{a,b}-1}{K_{a,b}} \right) f(\text{RB}_{a,b}) + \left( \frac{1}{K_{a,b}} \right) f(\text{RB}_{a,b} + 1) \right] \end{aligned}$$

där  $K_{a,o}$  nu betecknar storleken på den angivna besparingen enligt resonemanget i avsnitt 4.2 (d.v.s.  $K_{a,o}$  är här ett värde mindre än noll). Skillnaden i biasrisk kan i detta fall skrivas:

$$[\text{Biasrisk} - \text{Biasrisk}^-] = \text{Vtal}_a \cdot \text{Bet}_b \cdot \omega \cdot \left[ \frac{f(\text{RB}_{a,b}) - f(\text{RB}_{a,b}+1)}{K_{a,b}} \right] \quad (3)$$

Utifrån (1) – (3) kan nu skillnaderna i medelfel och biasrisk för varje tillförd eller besparad krona beräknas. För alla delundersökningar,  $u$ , och biaskomponenter,  $o$ , där riskbetyget angivits som ett värde större än 1 kommer en skillnad i biasrisk enligt (2) att beräknas. Om riskbetyget har satts till en etta och en negativ kostnad har angivits, så tolkas denna som ett besparingsförslag och skillnaden i biasrisk enligt (3) beräknas. För medelfelet beräknas skillnaden enligt (1).



Nu kan vi rangordna alla  $U \times O$  möjliga investeringar/besparingar efter den absoluta skillnad i medelfel eller bias som den skulle ge upphov, till för att se vilken som på marginalen ger mest kvalitet till KPI totalt.

## 5 Preliminära resultat (pilotstudie)

För att testa modellen har tre delundersökningar utvärderats preliminärt i en pilotstudie.

Pilotstudien indikerar att det tar runt en halv arbetsdag att fylla i mallen för en viss delundersökning. I takt med att mallen förenklas och vägledningarna blir mer detaljerade lär dock denna process gå snabbare. Notera även att det inte är nödvändigt, eller ens önskvärt, att samtliga delundersökningar utvärderas. Tanken är att börja med de delar av KPI där erfarenheten säger att det finns utrymme för förbättringar eller kostnadsbesparingar. När en undersökning väl värderats behöver denna värdering inte uppdateras förrän det skett betydande förändringar i undersökningens upplägg. Vägningstalen kan dock med fördel uppdateras kontinuerligt.

De undersökningar som utvärderades i pilotstudien är *Flygcharter*, *Räntor*, *Kläder* och *Utrikes flygresor*. Resultatet sammanfattas i tabell 1.

**Tabell 1:** Resultat av pilotstudie; de fem mest prioriterade åtgärderna.

| Delundersökning (u) | Komponent (o)   | Specifikation av åtgärd   |
|---------------------|---|---|
| Flygcharter         | Urval - Medelfel  | Öka urvalsstorleken   |
| Flygcharter         | Bearbetning –<br>Dataregistrering och<br>uppgiftslämnargranskning | Lägg över excelberäkningar till<br>SAS- och SQL-baserade<br>beräkningssystem                                |
| Utrikes flyg        | Urval - Medelfel  | Öka urvalsstorleken   |
| Kläder              | Bearbetning –<br>Dataregistrering och<br>uppgiftslämnargranskning | Förbättra kontrollerna i KPI mobile<br>för att förhindra att egenskaper och<br>prissignaler anges felaktigt |
| Kläder              | Ramtäckning –<br>Vägningstalens kvalitet                          | Undersök möjlighet till<br>transaktionsdata från ett antal större<br>klädkedjor                             |

## 6 Diskussion och frågor till nämnden

I denna PM har en möjlig resursallokeringsmodell för KPI presenterats.

Vi tänker oss att en förenkling av modellen är möjlig och skulle göra den mer användbar. Exempelvis är 26 kvalitetskomponenter nog mer än nödvändigt då det i vissa fall har visat sig vara svårt att veta till vilken komponent en viss felkälla tillhör. Målet är att använda pilotresultaten för att utvärdera den preliminära modellen och komma fram till lämpliga förbättringar.

En brist i modellen är att den inte tar hänsyn till de beroenden som finns mellan olika osäkerhetskomponenter och åtgärder. En åtgärd kan vara lösningen på flera olika biasrisker och när så är fallet har vi i pilotstudien valt att lägga denna totalkostnad på samtliga osäkerhetskällor. Det innebär att modellen måste användas iterativt i betydelsen att man tar fram en åtgärd i taget. Riskvärderingar och kostnader justeras mellan iterationerna för att ta hänsyn till de redan prioriterade åtgärderna.

En annan brist är att risker i en vidare betydelse inte beaktas i modellen. Säg exempelvis att externa data börjar användas i en viss delundersökning. Kostnadsmässigt ger det kanske en besparing och modellen kommer även att visa på minskat medelfel. En ny risk är emellertid att KPI nu är beroende av en extern dataleverantör för den reguljära publiceringen.

Nämnden välkomnas att inkomma med synpunkter på den förslagna modellen generellt samt på inriktningen av det fortsatta arbetet.

## Källor

- Andersson, C., Nilsson, P. och Carlsson, E. (2012). *Resultatrapport: Översyn av central priser i KPI*. Intern promemoria.
- Boskin, M.J. et al. (1996). *Toward a More Accurate Measure of the Cost of Living, Final Report from Advisory Commission*, U.S. Senate Finance Committee.
- Boskin, M.J. et al. (1998). *Consumer prices, the consumer price index, and the cost of living*, Journal of Economic Perception, 12, 3-26.
- Draper, R., och Schoultz, C. (2016). *A model for index auditing – the Swedish experience*. Presentation framförd vid Nordic SPPI Seminar, 19-20 May 2016, Copenhagen.
- Draper och Fridén (2017). *Index Auditing Model in Sweden*. Presentation framförd vid 32nd Voorburg Group Meeting, New Delhi, Indien, 2017.  
<http://voorburggroup.org/Documents/2017%20New%20Delhi/Papers/4013.pdf>.
- ILO (2004). *Consumer price index manual – theory and practice*.  
[http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/presentation/wcms\\_331153.pdf](http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/presentation/wcms_331153.pdf).
- Ribe, M. (2017). *Bias i KPI*, PM till Nämnden för KPI, sammanträde nr. 3 2017-10-25.
- SCB (2016a). *Känslighetsanalys i BNP*. Intern promemoria.
- SCB (2016b), *Kvalitet för den officiella statistiken – en handbok*.  
<https://www.scb.se/contentassets/8c5cc234fbae4e0ba731b1f9c8ea2042/kvalitet-for-den-officiella-statistiken.pdf>.

## Bilaga 1a: Skattning av omallokeringsfaktorer enligt projektet "Optimal allokering 2016"

Låt  $k_i$  beteckna de rörliga kostnader som är förknippade med urvalet i delundersökning  $i$  under en viss tidsperiod, exempelvis det senaste helåret. (Kostnader som inte påverkas av stickprovsstorleken betraktas som fasta kostnader och ska inte räknas med i  $k_i$ .) Den totala rörliga kostnaden betecknas  $K$ , så vi har  $K = \sum k_i$  där summan går över alla delundersökningar. Vägningstal för de olika delundersökningarna betecknar vi  $w_i$ .

Låt  $n_i$  vara en funktion av stickprovsstorlekarna i olika dimensioner (läge, tid och produkt) för delundersökningarna, sådan att följande villkor är uppfyllda:

- (i) Variansen är omvänt proportionell mot  $n_i$ ;  $V_i = a_i/n_i$
- (ii) Kostnaden är proportionell mot  $n_i$ ;  $k_i = b_i \cdot n_i$

Här är  $a_i$  och  $b_i$  oberoende av allokeringen och kan alltså betraktas som konstanter;  $a_i$  innehåller information om prisvariationen inom den aktuella produktgruppsgruppen och  $b_i$  om kostnaderna relaterade till mätningar inom gruppen. Tolkningen av  $n_i$  behöver inte vara densamma för (i) och (ii) men ett antagande vi gör här är att relationen mellan de två tolkningarna är densamma inom alla undersökningar. (Vi hade lika gärna kunnat skriva  $V_i = \frac{a_i}{n_i^{(1)}}$  och  $k_i = b_i \cdot n_i^{(2)}$ , där  $n_i^{(2)} = g \cdot n_i^{(1)}$  för alla  $i$ , och där  $g$  är en godtycklig konstant.)

Om vi ändrar allokeringen kommer  $V_i$  och  $k_i$  för de olika delundersökningarna att påverkas via  $n_i$ . Om  $V_i^*$  betecknar den nya variansen vi får med den nya allokeringen,  $k_i^*$  den nya kostnaden och  $n_i^*$  den nya stickprovsstorleken, så gäller då  $V_i^* = a_i/n_i^*$  och  $k_i^* = b_i \cdot n_i^*$ .

KPI totalt kan skrivas som ett vägt medelvärde av delindex,  $KPI = \sum w_i \cdot KPI_i$ . Variansen för totala KPI med den nya allokeringen kan skrivas  $V = \sum w_i^2 V_i^*$ . Problemet blir alltså att minimera (med avseende på  $k_i^*$ ) uttrycket för den totala variansen. Genom att plugga in sambandet  $V_i^* = \frac{V_i k_i}{k_i^*}$  och sedan minimera under bivillkoret  $K = \sum k_i^*$  får vi lösningen

$k_i^* \propto \frac{w_i \sqrt{V_i k_i}}{\sum w_i \sqrt{V_i k_i}}$ . Därmed erhålles det vi valt att kalla för omallokeringsfaktorerna,  $u_i$ :

$$u_i = \frac{k_i^*}{k_i} = \frac{K \cdot w_i \sqrt{V_i/k_i}}{\sum w_i \sqrt{V_i \cdot k_i}}$$

I Bilaga 1b listas omallokeringsfaktorerna i storleksordning. En faktor på 2 kan tolkas som att dubbelt så mycket resurser bör läggas på den aktuella delundersökningens urval, medan 0.5 betyder att resurserna kan halveras. Resultaten ska dock tolkas med viss försiktighet eftersom de bygger på förenklande antaganden samt att de inte tar hänsyn till andra osäkerhetskällor än urvalsfelet.

## Bilaga 1b: Omallokeringsfaktorer enligt projektet ”Optimal allokering 2016”

| Delundersökning / delsystem     | Omallokeringsfaktor ( $u_i$ ) | Vikt   |
|---------------------------------|-------------------------------|--------|
| Räntekostnader                  | 5,05                          | 42,40  |
| BANKTJÄNSTER                    | 2,80                          | 10,32  |
| FLYGCHARTER                     | 2,77                          | 13,21  |
| Lokalpriser (telefon)           | 2,62                          | 89,31  |
| Tv-, film- och musiktjänster    | 2,43                          | 4,90   |
| MOTIONSUTÖVNING                 | 1,90                          | 11,55  |
| EI                              | 1,82                          | 38,84  |
| BIOBILJETT                      | 1,74                          | 2,31   |
| INTRÄDESBILJETT, NÖJEN          | 1,67                          | 2,37   |
| LOKALTRAFIK                     | 1,63                          | 12,37  |
| Bilreparationer och reservdelar | 1,57                          | 16,47  |
| Husvagnar och husbilar          | 1,55                          | 2,43   |
| NYA BILAR                       | 1,54                          | 26,03  |
| BÖCKER                          | 1,41                          | 3,12   |
| Lokalpriser (besök)             | 1,22                          | 90,06  |
| Hyra, garage                    | 1,17                          | 117,10 |
| BILUTHYRNING                    | 1,11                          | 2,45   |
| MUSEER                          | 1,06                          | 0,62   |
| FASTA NÄTTJÄNSTER               | 1,05                          | 10,23  |
| UTRIKES FLYGRESOR               | 0,96                          | 5,90   |
| Kläder                          | 0,90                          | 44,16  |
| Frukt och grönt                 | 0,84                          | 10,93  |
| BILJETT, IDROTTSTÄVL.           | 0,82                          | 3,90   |
| FRÖ/LÖK                         | 0,77                          | 1,51   |
| Försäkringar                    | 0,77                          | 15,62  |
| Dagligvaror (kassaregisterdata) | 0,74                          | 144,49 |
| JÄRNVÄGSRESOR                   | 0,74                          | 3,98   |
| INRIKES FLYGRESOR               | 0,72                          | 2,51   |
| Skor                            | 0,65                          | 7,62   |
| BEGRAVNINGSKOSTNAD              | 0,61                          | 1,80   |
| Kommunala tjänster              | 0,60                          | 9,30   |
| BREV OCH PAKET                  | 0,59                          | 1,52   |
| BROAVGIFTER M.M.                | 0,59                          | 1,54   |
| Kött, ost                       | 0,56                          | 13,51  |
| KÖRSKOLEUTBILDNING              | 0,56                          | 1,66   |
| Inspelad musik                  | 0,55                          | 2,56   |
| BÅTRESOR, IN- OCH UTRIKES       | 0,48                          | 0,87   |
| MOTORCYKEL                      | 0,45                          | 1,17   |
| Dagstidningar                   | 0,45                          | 3,30   |

| <b>Delundersökning / delsystem</b> | <b>Omallokeringsfaktor (<math>u_i</math>)</b> | <b>Vikt</b> |
|------------------------------------|---|-------------|
| REPARATION/UNDERHÅLL AV BÅT        | 0,45  | 2,09        |
| Bensin                             | 0,41  | 34,07       |
| PRIVAT UTBILDNING                  | 0,38  | 1,81        |
| KONTROLLBESIKTNING                 | 0,37  | 0,79        |
| PARKERINGSavgifter                 | 0,37  | 2,54        |
| ANNONSKOSTNAD                      | 0,36  | 0,41        |
| TEATERBILJETT                      | 0,33  | 1,56        |
| LÅNGFÄRDSBUSS                      | 0,33  | 0,58        |
| TANDLÄKARARVODE                    | 0,33  | 9,75        |
| ELDNINGSOLJA,EGNAHEM               | 0,32  | 0,83        |
| STÄDNING                           | 0,30  | 1,62        |
| Populärtidskrifter                 | 0,28  | 2,17        |
| DATORTILLBEHÖR                     | 0,24  | 1,65        |
| Båtar                              | 0,23  | 1,41        |
| TAXI                               | 0,22  | 3,95        |
| HUSHÅLLSGAS, HY O BO LÄG           | 0,15  | 0,48        |
| Dagligvaror (övrigt)               | 0,13  | 3,27        |
| Avskrivningar och reparationer     | 0,09  | 0,24        |
| HYRA SOMMARPLATS FÖR BÅT           | 0,09  | 0,34        |

## Bilaga 2: Mall för utvärdering av biasrisker – Ett första utkast

### Del 1: Urval

| Område att bedöma                                      | Beskrivn. av nuläget | Vägledning   | Betyg | Betygs-motivering | Kostnad | Kostnads-motivering |
|--|----------------------|--|-------|-------------------|---------|---------------------|
| Urvals-metod<br>-<br>försäljnings-ställe               |                      | Sannolikhets-urval eller subjektivt urval? Cutoff? Bedöms urvalet vara representativt? |       |                   |         |                     |
| Urvals-metod<br>-<br>Produkt-erbjudande                |                      | Sannolikhets-urval eller subjektivt urval? Cutoff? Bedöms urvalet vara representativt? |       |                   |         |                     |
| Urvals-dragningens frekvens<br>-<br>försäljningsställe |                      |  |       |                   |         |                     |
| Urvals-dragningens frekvens<br>-<br>Produkt-erbjudande |                      |  |       |                   |         |                     |
| Urval av mät-tidpunkt under månaden                    |                      |  |       |                   |         |                     |
| Medelfel   |                      |  |       |                   |         |                     |

**Del 2: Ramtäckning**

| Område att bedöma                           | Beskrivn. av nuläget | Vägledning   | Betyg | Betygs-motivering | Kostnad | Kostnads-motivering |
|---|----------------------|--|-------|-------------------|---------|---------------------|
| Under-täckning<br>-<br>Försäljnings-ställen |                      | Försäljnings-ställen som finns med i mål-populationen men inte i ramen.  |       |                   |         |                     |
| Över-täckning<br>-<br>Försäljnings-ställen  |                      | Försäljnings-ställen som inte finns med i mål-populationen men som ändå finns i ramen.   |       |                   |         |                     |
| Under-täckning<br>-<br>Produkt-erbjudande   |                      | Produkter som finns med i mål-populationen men inte i ramen.   |       |                   |         |                     |
| Över-täckning<br>-<br>Produkt-erbjudande    |                      | Produkter som inte finns med i mål-populationen men som ändå finns i ramen.  |       |                   |         |                     |
| Vägnings-talets kvalitet                    |                      | Finns det något annat som kan påverka bilden av ramtäckning? Hur stor är risken för indexets tillförlitlighet beroende på detta? |       |                   |         |                     |

### Del 3: Mätning

| Område att bedöma      | Beskrivn. av nuläget | Vägledning  | Betyg | Betygs-motivering | Kostnad | Kostnads-motivering |
|------------------------|----------------------|---|-------|-------------------|---------|---------------------|
| Typ av pris            |                      | <p>Idealet är ett transaktionspris. Alternativ är exempelvis listpriser, timpriser eller modellberäknade priser.</p> <p>Används "unit value price"? Om unit value price, är produkterna tillräckligt homogena?</p>  |       |                   |         |                     |
| Insamlingsmetod        |                      | KPI mobile, telefon, KRD, internet.   |       |                   |         |                     |
| Manuella fel           |                      | <p>Hur stor är risken att den som samlar in priset gör fel?</p> <p>Är det lätt att hitta rätt pris? Är arbetssituationen sådan att misstag lätt händer?</p>   |       |                   |         |                     |
| Inliers                |                      | <p>Med inliers menas priser som felaktigt lämnas oförändrade över tid.</p> <p>Orsaken till detta är att det kan vara lätt för uppgiftslämnare att ange samma pris plus att det flyger under radarn vid vår granskning.</p>  |       |                   |         |                     |
| Produktspecifikationer |                      | <p>Med för bred specifikation avses att specifikationen inte är tillräckligt detaljerad och där vi riskerar att få in en kvalitetsförändring eller ändrad sammansättning av sålda produkter som prisförändring.</p> <p>Med för tigha specifikationer avses att vi har en allt för detaljerad beskrivning av produkt. Detta kan leda till att vi får svårt att hitta jämförbara transaktioner och har stort bortfall och många varubyten.</p> <p>Hur ofta uppdateras produktspecifikationerna?</p> <p>Konsument-profiler är en typ av specifikation.</p> |       |                   |         |                     |



**Del 4: Bortfall**

| Område att bedöma                           | Beskrivn. av nuläget | Vägledning   | Betyg | Betygs-motivering | Kostnad | Kostnads-motivering |
|---|----------------------|--|-------|-------------------|---------|---------------------|
| Objekts-bortfall på produkt-erbjudande-nivå |                      | Produkterbjudandet saknas  |       |                   |         |                     |
| Partiellt bortfall på produkterbjudandenivå |                      | Fattas information om produkterbjudandet? Ex, märket på en sko, typen av tyg i en kjol eller information om avbeställnings-skydd på en resa. |       |                   |         |                     |
| Bortfall på försäljnings-ställenivå         |                      | Uppgifter för ett helt försäljningsställe saknas.  |       |                   |         |                     |

**Del 5: Bearbetning**

| Område att bedöma                               | Beskrivn. av nuläget | Vägledning | Betygs-motivering | Betyg | Kostnad | Kostnads-motivering |
|---|----------------------|------------|-------------------|-------|---------|---------------------|
| Dataregistrering och uppgiftslämningsgranskning |                      |            |                   |       |         |                     |
| Mikrogranskning                                 |                      |            |                   |       |         |                     |
| Makrogranskning                                 |                      |            |                   |       |         |                     |

**Del 6: Modellantaganden**

| Område att bedöma         | Beskrivn. av nuläget | Vägledning  | Betyg | Betygs-motivering | Kostnad | Kostnads-motivering |
|---------------------------|----------------------|---|-------|-------------------|---------|---------------------|
| <b>Kvalitetsvärdering</b> |                      | Hur många byten? Vilken metod används för kvalitetsvärdering?                                       |       |                   |         |                     |
| <b>Cut off</b>            |                      |   |       |                   |         |                     |
| <b>Säsongsvaru-metod</b>  |                      | Innehåller produktgruppen säsongsvaor och i sådana fall, hur behandlas de? (Om inte, sätt betyg 1.) |       |                   |         |                     |
| <b>Indexformel</b>        |                      |   |       |                   |         |                     |