

Mätning av Skatteverkets effekter på dess omgivning

**Ett diskussionsunderlag och hjälpmedel
för såväl beslutsfattare som utförare
av effektmätningar**

Rapport 2008:2

Förord

På uppdrag av produktionsavdelningens ledningsgrupp har analysenheten tagit fram ett diskussionsunderlag och en handledning kring hur man kan mäta effekter av Skatteverkets åtgärder på dess omgivning.

Rapporten behandlar effektmätning ur flera perspektiv och utmynnar i bl.a. förslag på kriterier som bör vara uppfyllda för att det ska vara aktuellt att mäta effekter av Skatteverkets åtgärder på dess omgivning. Dessutom görs en analys av förhållandet mellan Skatteverkets övergripande mål och effekter av olika åtgärder, samt introduceras en modell för att kartlägga skatteuppbörden i relation till grad av systemutbyggnad, kallad medverkandekartan. På detta sätt utgör delar av rapporten ett komplement till den nyligen lanserade skattefelskartan.

Ansvariga för framtagande av rapporten har varit Mats Andersson, Joacim Danielsson och Mikael Pyka på Skatteverkets analysenhet. Författarna svarar själva för innehållet i rapporten.

Solna i april 2008

Jan-Erik Bäckman
/Chef för analysenheten/

Innehållsförteckning

Sammanfattning	7
Del I: Introduktion	17
1 Inledande läsanvisning	17
2 Varför är det intressant att mäta effekter?	17
2.1 Effekter, resultat, prestationer – vad är det för skillnad?	18
3 Att fundera över effekter	19
3.1 Verksamhetslogik.....	20
3.2 När inträffar effekten, och hur länge varar den?	20
3.3 Attributionsproblemet	21
3.4 Evidensdiskussionen	21
3.5 Ramar för effektmätning	22
3.5.1 Kostnad och tid.....	22
3.5.2 Etik	23
3.6 Vem ska ansvara för effektmätning?.....	23
3.7 Fördelar med effektmätning	24
3.8 Risker med effektmätning – vad säger kritikerna?.....	24
3.8.1 Osäkra mätningar	24
3.8.2 Målförskjutning.....	25
3.8.3 Andra exempel på oönskade bieffekter – avprofessionalisering och känsla av övervakning.....	26
3.9 Konflikt mellan det väsentliga och det mätbara?.....	26
Del II: Effektmätning utifrån Skatteverkets övergripande mål	29
4 Skatteverkets resultatuppföljningsmodeller och en idealisk effektmätningmodell	29
4.1 Resumés bild av verksamheten och uppföljningen av den	29
4.2 Guppis bild av verksamheten och uppföljningen av den	30
4.3 De viktigaste skillnaderna mellan Resumé och Guppi.....	31
4.4 Sammanfattning och slutsatser kapitel 4.....	31
5 Strategiskt perspektiv - vad kan mätas med utgångspunkt i våra övergripande mål? 33	33
5.1 Övergripande mål	33
5.1.1 Formulering av övergripande effektmål utifrån regleringsbrevet	33
5.1.2 Medborgarens uppdrag till Skatteverket - rättighetssidan.....	34
5.1.3 Medborgarens skyldigheter	35
5.1.4 Måluppfyllelsen kan mätas med både subjektiva och objektiva mått.....	35
5.1.5 Det övergripande målet uppdelat i samtliga dimensioner	36
5.1.6 Andra politikområden	40
5.2 En tidsindelning av aktiviteterna kan vara utgångspunkt för effektmätning	41
5.2.1 Medverkandekartan	43
5.3 Vad som kan mätas av enskilda åtgärder – sammankoppling av åtgärder och måldimensioner	46
5.3.1 Den första analysen - uppföljning av åtgärder och mätning av effekter	47
5.3.2 Utveckling av den första analysen – följd effekter visar sig i ett längre perspektiv	49
5.3.3 Den andra analysen – hur kan modellen användas för att hitta förbättringsmöjligheter i verksamheten?.....	51
5.3.4 Utveckling av den andra analysen - ett cykliskt betraktelsesätt av åtgärdstyperna kan ge ytterligare information och kunskap.....	53
5.4 Sammanfattning och slutsats kapitel 5	54
6 Operativt perspektiv - vad kan mätas om utgångspunkten är enskilda insatser?	55
6.1 Kausaliteten.....	55

6.1.1 Olika insatsers påverkan i tiden	55
6.1.2 Går det att teoretiskt fastställa ett samband mellan en kontrollinsats och det långsiktiga målet?.....	56
6.2 Målet och dess mätbarhet	60
6.2.1 Validitet och reliabilitet.....	60
6.2.2 Mätbarhet genom operationalisering av effektmålet.....	61
6.2.3 Mätbarhet genom indikatorer som ligger i linje med effektmålet.....	63
6.2.4 Externa faktorer och reliabiliteten.....	64
6.2.5 Interna faktorer och målförskjutning.....	65
6.3 Sammanfattning och slutsats kapitel 6	65
Del III: Metoden för effektmätning.....	67
7 Inledning: Effekter och modeller för att beskriva insatsen.....	67
7.1 Effekter.....	67
7.2 Verksamhetslogik som hjälpmedel vid effektmätningar.....	68
7.2.1 Hur kan verksamhetslogiken användas i praktiken?.....	69
8. Att mäta effekter – kvalitetskriterier för en effektmätning	71
8.1 Kvaliteten på effektmåtten	72
8.1.1 Bestäm utifrån effektkejdan vad mätningen ska omfatta	72
8.1.2 Olika metoder för datainsamling	72
8.1.3 Teoretisk definition	73
8.1.4 Operationella definitioner: effektmått	73
8.2 Att isolera effekter.....	74
8.2.1 Introduktion: Isolerbarhetskriterierna och det ideala upplägget	74
8.2.2 Klassificering av upplägg för att isolera effekterna	76
8.2.3 Slumpmässigt kontrollerade experiment	77
8.2.4 Kvasi-experimentell ansats	80
8.2.5 Icke-experimentella metoder	81
8.3 Generaliserbarhet: för vilka är resultatet giltigt?.....	87
9. Att mäta effekter – utifrån en skatteförvaltnings förutsättningar	88
9.1 Olika upplägg möjliga för olika kategorier av insatser	88
9.2 Möjliga upplägg utifrån informationstillgång	89
9.2.1 Information av god kvalitet finns för hela populationen.....	89
9.2.2 Information av god kvalitet kräver mätning genom insats (kontroll)	90
9.2.3 Information kan tillgängliggöras för ett urval av populationen	91
10. Att belysa effekterna av en insats med kvalitativ metod	92
11. Avslutande diskussion.....	94
11.1 För vilka insatser är det meningsfullt att mäta effekter?.....	94
11.2 För vilka insatser är det värdefullt att lägga extra resurser för att kunna mäta dess effekter?.....	95
11.3 Vad gör man med resultatet av en effektmätning?.....	97
Litteraturlista.....	99
Bilaga. En fördjupning i metoden för effektmätning.....	103
Slumpmässigt kontrollerade experiment.....	103
Att estimeras effekter vid icke-experimentella studier	104

Sammanfattning

Del I: Introduktion till effektmätning

I den inledande delen av rapporten introducerar vi bakgrunden till att behovet av effektmätningar av offentlig verksamhet har uppstått. Effekter definieras som de förändringar i Skatteverkets omgivning som inträder på grund av olika åtgärder som Skatteverket har vidtagit, t.ex. förändrade beteenden hos olika grupper av skattebetalare.

Dessutom presenteras en enkel modell, här kallad utbyggd verksamhetslogik, att använda såväl när man planerar vilken insats man ska utföra för att uppnå ett givet effektmål, som när man ska utvärdera effekterna av redan utförda insatser. Det är viktigt att poängtera att planering av olika åtgärder även bör innehålla planering för hur effekterna ska mätas (om de är mätbara).

Att ta sikte på effekter kan skapa kreativitet och lärande inom organisationen. Men detta förutsätter att man samtidigt är medveten om dels risken för mätproblem i form av osäkra mått och höga kostnader, dels risken för målförskjutning om man inom organisationen känner sig tvingad att ersätta ett effektmått med ett mått på prestationer för att uppfylla ett krav på mätbara effektmått. Produktivitet och kvalitet behöver givetvis mätas, men är inte samma sak som att mäta effekter. I de fall ansvariga för en åtgärd efter utredning bedömer att effekterna inte går att mäta bör detta motiveras i särskild ordning. Därmed ansvarar dessa på ett tydligt sätt även för denna bedömning.

Del II: Effektmätning utifrån Skatteverkets övergripande mål

Denna del av rapporten behandlar effektmätning ur ett myndighetsperspektiv. Den utgår från Skatteverkets övergripande mål och består av tre underavdelningar. Den första avdelningen (kapitel 4) behandlar hur effektmätning betraktas i de två verksamhetsuppföljningsmodeller som under senare år använts inom Skatteverket: Resumé och Guppi. Den andra avdelningen (kapitel 5) har ett strategiskt effektmätningssperspektiv. Den tar sin utgångspunkt i de övergripande målen och går från dessa bakåt mot olika insatser, för att den vägen undersöka vilka kopplingar som finns mellan insatser och effekter. Den tredje avdelningen (kapitel 6) går den motsatta vägen. Den tar sin utgångspunkt i enskilda insatser och går från dessa framåt mot de övergripande målen för att den vägen se om det går att uttala sig om eventuella effekter. Den sista delen har ett mer operativt perspektiv. Nedan följer en kort sammanfattning av respektive kapitel.

Kapitel 4

Skatteverket släppte år 1997 en rapport som beskrev verksamhetsuppföljningsmodellen Resumé. Denna modell tog sin utgångspunkt i de övergripande målen. De övergripande målen bröts sedan ner i resultatkriterier på tre nivåer. Den lägsta mätbara nivån utgjordes av resultatindikatorer bestående av olika slag av prestationer, t.ex. ändringsfrekvenser.

Prestationer har som mått betraktat den fördelen att de är lätta att mäta. En nackdel med att låta den typen av mått styra verksamheten kan dock vara att de eftersträfvade effekterna inte uppnås. Verksamheten riskerar med ett strikt prestationsfokus att styras fel. För att i högre grad uppnå effektstyrning av verksamheten utarbetades under 2004 en ny verksamhetsuppföljningsmodell kallad Guppi. Denna modell utgår från det övergripande målet att minimera skattefelet, som står under restriktionen att skattebetalarna ska göra rätt från början.

Målet uppnås genom en hel uppsättning faktorer och åtgärder, såsom att skapa förtroende för skattesystemet och Skatteverket, öka den upplevda risken för upptäckt och normpåverkan.

Guppis sammanslagning av åtgärder och påverkansfaktorer, till hela enheter, medför ur effektmätningssynpunkt ett behov av en utveckling av modellen. Effektmätning handlar om enskilda insatsers påverkan på det övergripande målet. En användbar effektmätningmodell måste därför arbeta vidare utifrån Guppi, och strukturera upp såväl övergripande mål som enskilda insatser. På så sätt kan sedan möjligheter uppstå att finna kopplingarna mellan enskilda insatser och övergripande mål, vilket är vad denna del fortsättningsvis strävar mot.

Kapitel 5

Detta kapitel har ett strategiskt perspektiv och tar sin utgångspunkt i Skatteverkets övergripande mål så som det formuleras i regleringsbrevet.

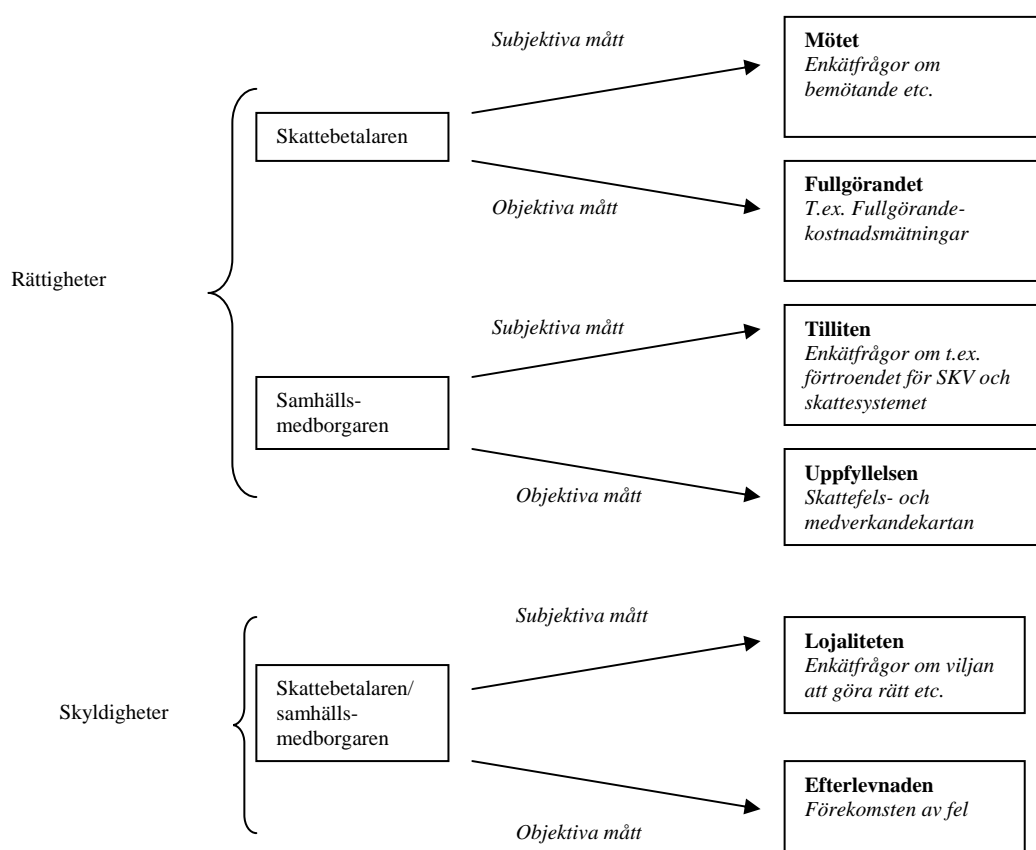
När regleringsbrevets mål struktureras upp går det att identifiera två uppdrag från den enskilde medborgaren till Skatteverket. Det första uppdraget är direkt och handlar om att den enskilde medborgaren ska kunna fullgöra sina skyldigheter som skattebetalare på ett så enkelt, smidigt och billigt sätt som möjligt. Detta uppdrag ges som *skattebetalare*. Det andra uppdraget handlar om att Skatteverket ska ta in den skatt som är demokratiskt beslutad. Den enskilde ska alltså kunna lita på att Skatteverket fullgör sin uppgift att säkerställa finansieringen av de gemensamma nyttigheterna på ett så bra sätt som möjligt. Detta andra uppdrag ges som *samhällsmedborgare*.

Denna del är medborgarens rättighetssida, där den enskilde medborgaren i egenskap av dels skattebetalare, dels samhällsmedborgare har vissa krav på Skatteverket. Det finns dock även en skyldighetssida. Den handlar om att den enskilde är skyldig att själv betala sin skatt i den omfattning som lagarna föreskriver.

Det övergripande målet uppdelat på detta sätt kan sedan mätas med olika metoder. Två grundläggande sätt är att antingen mäta med subjektiva mått - attityder eller föreställningar om något - eller med objektiva mått - t.ex. kronor och tidsenheter.

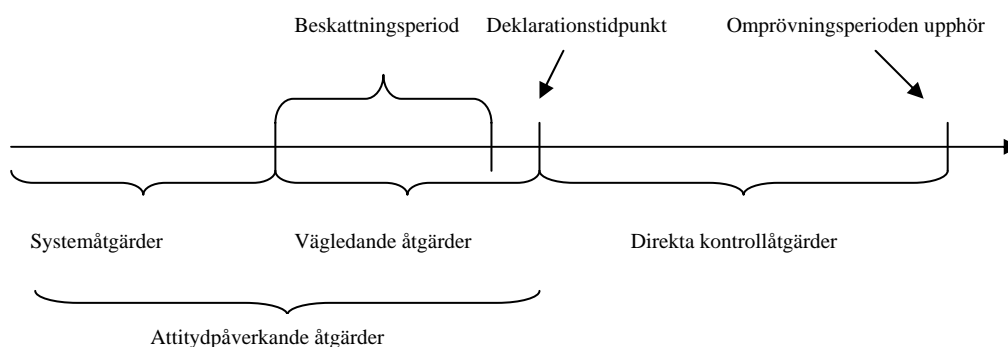
Det övergripande målet uppdelat som ovan, med de två olika grundläggande sätten att mäta, ger sex olika dimensioner. Varje sådan måldimension har fått en benämning. Dessa benämningar är *Mötet*, *Fullgörandet*, *Tilliten*, *Uppfyllelsen*, *Lojaliteten* och *Efterlevnaden*. Vilken dimension av det övergripande målet som de avser att mäta framgår av illustrationen nedan.

Fig II.1 Det övergripande målets olika dimensioner



När det övergripande målet är uppdelat på detta sätt går kapitlet vidare till de olika insatser eller åtgärder som Skatteverket har tillgång till. De olika insatserna kategoriseras utifrån vilka möjligheter som finns att påverka beskattningen för en *enskild beskattningsperiod* eller ett *enskilt beskattningsår*. Detta kan illustreras som i tidslinjen nedan.

Fig II.2 De olika åtgärdstypernas påverkan i tiden



Till och med beskattningsperiodens början så kan beskattningen påverkas genom olika typer av systemåtgärder. Med systemåtgärder menas allt ifrån lagstiftningsändringar till blankettförändringar. Fram till deklarationstidpunkten kan skattebetalarna påverkas genom antingen vägledande åtgärder, som hjälper dem att göra rätt, eller attitydpåverkande åtgärder, som får dem att vilja göra rätt. Till attitydpåverkande åtgärder räknas den preventiva effekten

av tidigare perioders kontroller. Vägledande åtgärder är definitionsmässigt mest inriktade mot de lojala skattebetalarna, medan de attitydpåverkande åtgärderna främst har betydelse för de illojala. När deklarationstidpunkten har inträffat för den aktuella beskattningsperioden är det försent att göra rätt från början. Skatteverket kan från denna tidpunkt endast påverka skattefelet genom direkta kontrollåtgärder. De direkta kontrollåtgärderna har sedan en effekt på skattefelet till och med omprövningsperiodens slut. Då blir skattefelet avseende det aktuella beskattningsåret definitivt.

Denna tidsindelning av åtgärderna leder till åtgärdshierarkin nedan.

- Systemåtgärder
- Vägledande åtgärder
- Attitydpåverkande åtgärder
- Direkta kontrollåtgärder

Denna åtgärdshierarki bygger på synsättet att det billigaste, smidigaste och mest effektiva sättet att beskatta är genom väl utbyggda system för redovisning, fastställande och kontroll. Inom de områden som inte helt täcks av systemen kan vägledande åtgärder utgöra ett komplement till de skattebetalare som vill göra rätt men finner det komplicerat, t.ex på grund av svårbegripliga regler. Inom områden som inte täcks av effektiva system, och där vägledning inte har någon effekt, kan Skatteverket gå in med attitydpåverkande åtgärder för att öka viljan att göra rätt för sig. När redovisningstidpunkten har inträffat är alla dessa möjligheter uttömda. Det som återstår är därmed kontroller, som är den sista möjligheten att påverka skattefelet för det aktuella året.

Tidsindelningen ovan leder även fram till vad som kallas medverkandekartan. Den bygger på följande resonemang. Olika skatter fastställs i olika grad med hjälp av system. Tjänsteinkomster och fastighetsskatt är skatter som i mycket hög grad fastställs genom system. Kapitalskatter fastställs i ganska hög grad genom kontrolluppgifter, och därmed genom system. Kan de inte fastställas helt så finns ändå i många fall goda kontrollindikationer. För arbetsgivaravgifter och moms finns mekanismer som i viss grad stödjer skattekontrollen. För bolagsskatten finns andra faktorer som i någon mån stabiliserar skattebasen. Andra skatter, såsom de som utgår på enskild näringsverksamhet, bygger i mycket hög grad på skattebetalarnas egna aktiva medverkan.

Medverkandekartans idé är att kategorisera den totala skatteuppbörden för ett visst år efter i vilken grad den säkerställts genom systemen. Detta skulle innebära att basen i den så kallade skattestapelns strukturerades upp enligt denna princip och att medverkandekartan på så sätt utgör ett komplement till skattefelskartan, som koncentrerar sig på stapelns övre del.

Medverkandekartan skulle därmed även utgöra ett av de två viktigaste måtten på den del av det övergripande målet som kallas *Uppfyllelsen*, dvs hur Skatteverket objektivt sköter uppdraget från samhällsmedborgarna. Det andra måttet på *Uppfyllelsen* är skattefelskartan.

Kapitlet avslutas med en sammankoppling av de fyra åtgärdskategorierna och de sex måldimensionerna, för ett resonemang om vilka samband som finns mellan dem och vilka möjligheter modellen ger för strategiska analyser.

Kapitel 6

Detta kapitel har ett mer operativt perspektiv och tar sin utgångspunkt i de enskilda insatserna för att sedan undersöka vad som kan sägas om deras eventuella effekt på de övergripande

målen. Den första delen undersöker kausaliteten mellan insatser och övergripande mål, uttryckt som en minskning av skattefelet. Den andra delen behandlar målformulering.

I den första delen konstateras först att Skatteverkets strategi att skattebetalarna ska förmås att göra rätt från början är komplicerad ur effektmätningsspektiv. Av de fyra åtgärdsstyper som Skatteverket har att laborera med - systemåtgärder, vägledning, attitydpåverkan och kontroll - har de tre första en effekt innevarande år medan den sista i bästa fall har en effekt efterföljande år. De tre första åtgärderna utförs till och med deklarationstidpunkten. Efter denna är det försent att göra rätt från början. Kontrollen kan genomföras först efter deklarationstidpunkten, och kan därmed i bästa fall leda till att skattebetalarna gör rätt i nästa beskattningsperiod. Vidare konstateras att det finns en mängd faktorer som påverkar kontrollens effekter på skattefelet, och att det är långtifrån säkert att kontroller alltid leder till ett minskat skattefel genom preventiv effekt.

I den andra delen behandlas mätning av enskilda insatsers effekter på det övergripande målet genom indikatorer. I denna del belyses olika problem och risker med detta. Exempel på sådana är att irrelevanta delmål eller indikatorer, som endast skenbart mäter det man vill mäta, kan ge felaktig information eller till och med leda verksamheten fel.

Slutsatsen av detta kapitel är att det inte finns en enda enkel modell som kan användas för mätning av effekter av alla olika typer av insatser. Istället måste varje ny effektmätning genomföras utifrån sina egna förutsättningar. Planering och upplägg blir ofta, till större eller mindre del, unik för varje ny effektmätning som ska utföras. I nästa del presenteras praktiska riktlinjer för detta.

Del III: Metoden för effektmätning

Denna del är mer praktiskt inriktad och syftet är att på ett överskådligt sätt beskriva hur effektmätning av Skatteverkets insatser kan genomföras.

Del tre kan sammanfattas i form av ett praktiskt hjälpmedel, *flödesschema för effektmätning*, som är utformat för att vägleda beställaren av insatsen och utföraren av effektmätningen redan på insatsens planeringsstadium (se nedan). Den pekar på effektmätningens olika steg och beslutspunkter. Utgångspunkten är att se om det finns naturliga förutsättningar att mäta effekterna av god kvalitet, eller om det skulle krävas t.ex. att insatsen utformades på ett särskilt sätt för att detta ska vara möjligt. Observera dock att flödesschemat med nödvändighet innebär en förenkling av förutsättningarna för effektmätning.¹

Vi börjar dock med att sammanfatta huvudtankarna i del 3.

Tre kvalitetskriterier

Vi identifierar tre kriterier som avgör kvaliteten på effektmätningen:

- Kvaliteten på effektmåtten. Detta kriterium rymmer två frågor:
 - I vilken mån återspeglar effektmåtten de verkliga effekterna, och
 - i vilken mån är mätningen pålitlig?
- Isolerbarhet: I vilken mån är den eventuellt observerade förändringen av effektmåtten över tiden (före och efter insatsen) ett resultat av själva insatsen?
- Generaliserbarhet: För vilka är resultatet giltigt?

¹ Förenklingen innebär t.ex. att ett av kvalitetskriterierna (generaliserbarheten) inte ingår, ej heller de mer avancerade icke-experimentella metoderna (såsom metoder för efterhandsmatchning). Informationskategori 1 och 3 hanteras också i samma flöde, vilket är lite missvisande (se avsnitt 9.2.).

Kvaliteten på effektmåtten

Effektmåtten ska alltså väl beskriva de verkliga effekterna. Som effektmått för att utvärdera effekterna av införandet av personalliggarna används bl.a. redovisade löner hos berörda företag. Detta svarar väl mot effekter av minskat svartarbete man önskar uppnå. Tillförlitligheten i mätningen är perfekt i den meningen att uppgifterna är de som företagen själva har lämnat.

Isolerbarhet

Att isolera effekterna av en insats innebär i allmänhet en stor utmaning, då en mängd andra föränderliga faktorer kan påverka utfallet. Utmaningen består i att beskriva hur tillståndet skulle ha sett ut om insatsen inte hade genomförts, det s.k. kontrafaktiska tillståndet.

Den typ av insatser som är enklast att effektmäta är de som uppfyller det vi kallar *isolerbarhetskriterierna*. De beskriver en situation då insatsen är den huvudsakliga påverkansfaktorn, så att en ev. förändring i stor utsträckning kan tillskrivas insatsen i fråga:

- Tiden för själva insatsen är kort
- Effekten inträffar direkt efter insatsen
- Tiden mellan mätpunkterna är kort
- Insatsen är omfattande i förhållande till förändringarna i omgivningen

Med det sista kriteriet menas att andra påverkansfaktorer i jämförelse med insatsen är små och/eller förändras lite över tiden. Om dessa kriterier är uppfyllda ger en jämförelse mellan tillståndet före respektive efter insatsen en god bild av insatsens effekter. Exempel på insatser som uppfyller dessa villkor kan vara genomgripande lagstiftningsförändringar eller då nya e-lösningar introduceras.

Det ideala upplägget för att isolera effekterna av en insats är att från populationen slumpmässigt välja ut två grupper, och låta endast den ena gruppen bli föremål för insatsen. Skillnaden i utfallet mellan grupperna motsvaras av effekten av insatsen.

Ovan nämnda experimentella upplägg, och varianter som är bättre förenliga med en effektiv verksamhet, kan under vissa förutsättningar användas för att testa nya arbetssätt och metoder, till en låg merkostnad. Experiment kan dock vara associerade med etiska problem och högre kostnader.

För vissa insatser går det att identifiera ”naturliga” referensgrupper som väl beskriver det kontrafaktiska tillståndet, t.ex.:

- Liknande grupp, samma tid
- Samma grupp, annan/”liknande” tid
- Liknande grupp, ”liknande” tid

Om informationstillgången är sådan att det går att följa effektmåttets utveckling över tiden kan tidsseriebaserade analyser göras, t.ex. kan utvecklingen jämföras med utvecklingen i en referensgrupp eller s.k. tidsserieanalys användas.

Regression discontinuity design är det krångliga namnet på en relativt enkel, men elegant, metod som är lämplig att använda för att uppskatta de individualpreventiva effekterna av Skatteverkets utökade kontroller, inom ett antal områden.

I bilagan introduceras även tämligen avancerade icke-experimentella metoder för efterhandsmatchning m.m. som kan vara intressanta att använda för mer omfattande och strategiskt viktiga insatser.

Generaliserbarhet

För insatser där urval görs är resultatet av effektmätningen giltigt för de som omfattas av insatsen. Endast i de fall då urvalet är slumpmässigt är resultatet giltigt för hela populationen.

Effektmätning av Skatteverkets insatser – särskilda förutsättningar

Förutsättningarna för effektmätning av de insatser som Skatteverket genomför beror dels på vilken typ av insats det handlar om, dels informationstillgången.

Störst möjligheter att *välja* ett upplägg som kan isolera effekterna av en insats har vi då ett urval av en grupp skattebetalare blir föremål för insatsen (t.ex. skrivbordskontroll eller riktade informationsinsatser). Minst möjligheter finns då hela populationen blir föremål för insatsen (t.ex. vid en lagstiftnings- eller blankettförändring).

Informationstillgången för en insats tillhör någon av följande kategorier:

1. Information erhålls med god kvalitet, vilken väl återspeglar de verkliga effekterna, för samtliga i populationen.
2. Information av god kvalitet, efter genomförd kontroll (som insats), kräver också mätning genom kontroll.
3. Information av god kvalitet erhålls ej för samtliga i populationen, men information som svarar väl mot de verkliga effekterna kan inhämtas utan att mätning och insats sammanfaller.

För kategori 1 innebär informationstillgången ingen begränsning av möjliga upplägg. Kostnaden för mätningen är också relativt liten. För insatser som har en informationstillgång enligt kategori 2 är en effektmätning, i normalfallet, dyr och etiskt tveksam. För kategori 3 innebär informationstillgången, i allmänhet, ingen begränsning av möjliga upplägg. Mätningen riskerar dock att bli dyr och mätproblem uppstår om mätningen i sig påverkar skattebetalarna.

Kvalitativ metod

Om vi söker svaret på frågan om *varför* en åtgärd har eller inte har effekt, bör man tillämpa någon form av kvalitativ metod. Detta kan t.ex. innefatta att intervjua de som utsatts för en åtgärd, intresseorganisationer för dem som utsatts för åtgärden eller någon grupp av experter på området. Intervjuer kan utföras en och en, eller i olika former av gruppkonstellationer. Denna typ av undersökningar lämpar sig ofta för någon form av konsulter att utföra, eftersom man i vissa situationer kan förvänta sig en problematisk intervjuareffekt av att Skatteverket försöker få svar på denna typ av frågor. Andra exempel på kvalitativa metoder är benchmarks, där man försöker jämföra våra erfarenheter med t.ex. skatteförvaltningar i andra länder.

Avslutande diskussion

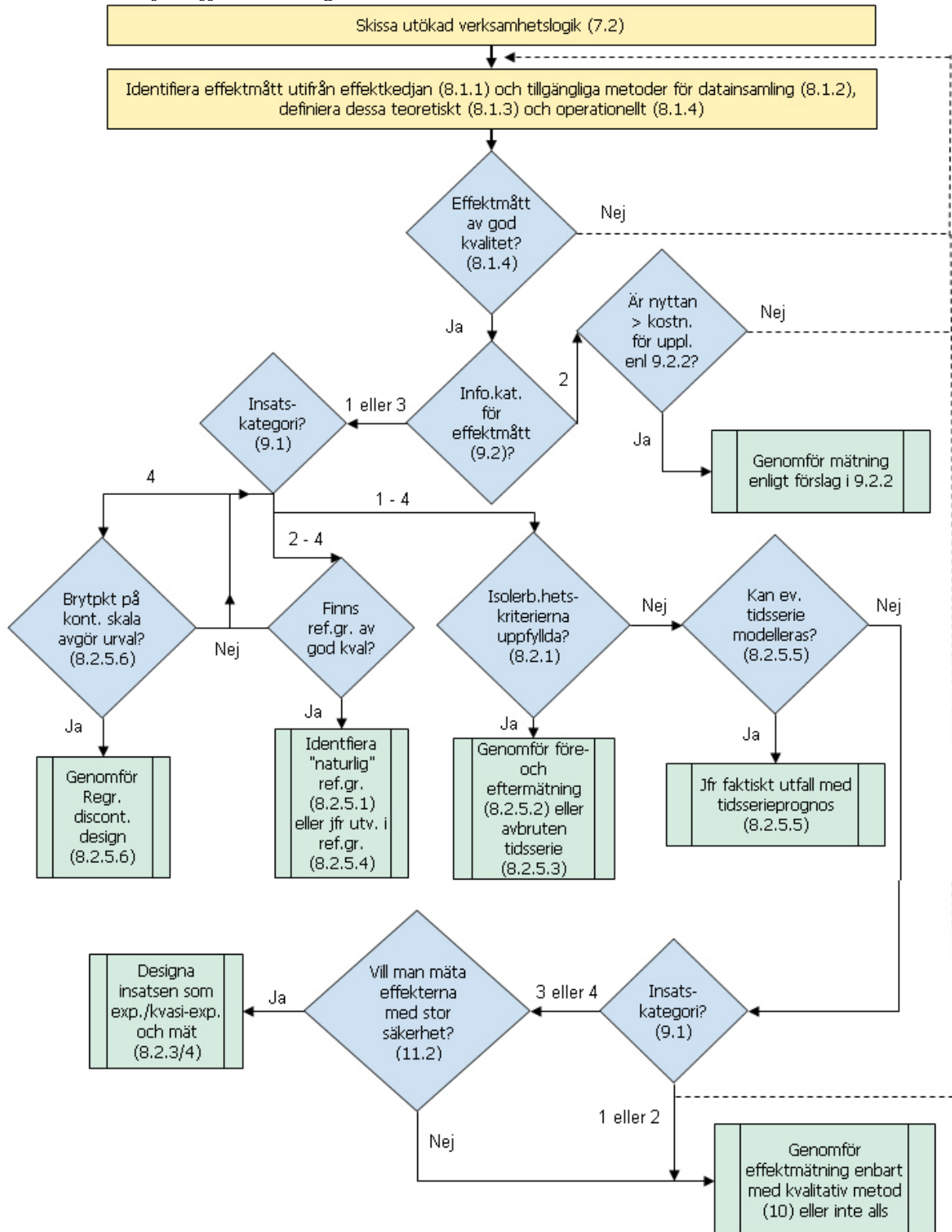
Vår bedömning är att det är meningsfullt att försöka *mäta* effekter då kvaliteten på effektmåtten är god eller ganska god *och* det går väl eller ganska väl att isolera effekterna av insatsen, men inte annars. I de fall då dessa förutsättningar inte ”naturligt” är uppfyllda, kan det vara möjligt att tillskapa dessa förutsättningar genom något experimentupplägg och/eller extra informationsinhämtning. I dessa fall kan det krävas betydande resurser för att möjliggöra en effektmätning av god kvalitet, varför vi rekommenderar att följande frågor beaktas:

1. Vilken nytta skulle det innebära att mäta effekterna av insatsen med god kvalitet?
2. Vilket mervärde innebär det att, om möjligt, välja ett upplägg som bättre kan isolera effekterna av insatsen? Vilka merkostnader innebär detta/dessa uppläggen?

3. Vilken grad av kvalitet och kostnad är de olika alternativen för informationsinhämtning associerade med?

I rapportens avslutande avsnitt exemplifierar vi hur resultatet av en genomförd effektmätning kan tolkas utifrån den utökade verksamhetslogiken och ligga till grund för ett kontinuerligt lärande.

Flödesschema för effektmätning



Förklaring till flödesschemat

- För att beskriva vilka effekter man hoppas uppnå med insatsen, och varför, är en *effektkedja* en lämplig modell. Börja med att skissa den. Lägg därtill de *externa faktorer* som också kan påverka utfallet och s.k. *ramfaktorer* som är möjliggörare och ev. risker/hinder för att insatsen ska uppnå sina effekter. (Se 7.2)
- Identifiera utifrån effektkedjan om mätningen kan omfatta en eller flera långsiktiga effekter, indikatorer på de långsiktiga effekterna, eller deffekter (8.1). Definiera dessa effektmåts innebörd (8.1.3) och hur de ska mätas (8.1.4). Viktigt att se till *före* insatsen påbörjas är att relevant information finns tillgänglig antingen i t.ex. Skatteverkets register eller inhämtas separat, helst före, under och efter insatsen. (8.1.2)
- Relevant information kan vara kostsam att samla in och då måste värdet av denna information vägas mot ev. billigare information som mindre väl återspeglar de verkliga effekterna. Frågan är om det finns effektmått av god kvalitet, till en rimlig kostnad. (8.1.4)
- Därefter måste vi fråga oss till vilken kategori (se ovan) informationen vi identifierat i föregående steg tillhör.
 - Om kategori 2: studera möjliga upplägg (9.2.2) och bedöm om nyttan med mätningen är större än kostnaderna (ekonomiska, etiska, m.m.) för densamma.
 - Om kategori 1 eller 3: bestäm till vilken kategori insatsen hör (se 9.1).
- Tillhör insatsen kategori 4 och det är en brytpunkt på en kontinuerlig skala som avgör vilka som blir föremål för insatsen eller ej? Om ja: mät i så fall effekterna med hjälp av metoden Regression discontinuity design (8.2.5.6).
- Om nej: Tillhör insatsen kategori 2 - 4 och det är möjligt att identifiera en referensgrupp av god kvalitet? Om ja: mät i så fall effekterna genom att jämföra med en eller flera "naturliga" referensgrupper (8.2.5.1) eller jämför utveckling av effektmåttet över tiden med utvecklingen av detsamma i referensgruppen (8.2.5.4).
- Om nej: Är isolerbarhetskriterierna väl eller ganska väl uppfyllda? Om ja: Genomför en före- och eftermätning (8.2.5.2) eller uppskatta effekterna med hjälp av en s.k. avbruten tidsserie (8.2.5.3).
- Om nej: Om effektmåttet kan följas i form av en tidsserie och det är möjligt att identifiera faktorer som delvis påverkar utvecklingen av effektmåttet, kan en s.k. tidsserieanalys användas för beskriva den kontrafaktiska situationen (8.2.5.5).
- Om ovanstående metoder inte kunnat användas återstår frågan om kunskapsbehovet rörande insatsens effekter är så stort att det motiverar en högre kostnad för mätning. De ovan nämnda tre frågorna rekommenderas vara vägledande för detta beslut (11.2).
 - Om insatsen ska genomföras som ett experiment, studera 8.2.3/4 och bilagan för olika alternativa upplägg. Använd ett upplägg som är väl anpassat för insatsen, planera och förankra det i god tid, samt förbered för rapportering av eventuella avvikelser från planen.
- Använd gärna kvalitativ metod för att belysa effekterna och tolka resultatet utifrån den uppställda verksamhetslogiken (10 och 11.3).
- Oavsett val av metod för effektmätning, förbered för hur resultatet av effektmätningen och dess slutsatser ska tas om hand och implementeras i verksamheten.

Del I: Introduktion

Under kommunismens tid i Polen mättes resultatet för möbelfabriker i enheten ”producerat antal ton möbler”. Som ett resultat av detta har Polen idag de tyngsta möblerna i världen. (citat ur Perrin, 1998, vår översättning)

1 Inledande läsanvisning

Denna skrift avser att beskriva hur man kan mäta effekter av Skatteverkets verksamhet. Avsikten har varit att beskriva vårt synsätt på effektmätning, givet vår bakgrund som utredare på Skatteverkets analysenhet och de frågor som ställts till oss i den effektmätningssgrupp som bildats för att utgöra metodstöd i effektmätning inom Skatteverket. Därutöver introducerar vi en modell, kallad medverkandekartan, som avses att utgöra ett komplement till övriga sätt att mäta effekter av våra insatser, t.ex. förtroendemätningar och skattefelskartan. Utifrån dessa perspektiv har vi tagit vissa saker för givna, t.ex. produktivitets- och kvalitetsmätningar, åiterrapporteringskrav i årsredovisningen etc. som vi inte berör i denna rapport.

I detta inledande kapitlet introducerar vi teorierna om varför effektmätning över huvud taget är intressant att utföra – vilka fördelar som det medför samt vilka kända risker som finns med detta styrmedel. I del II diskuterar vi förhållanden mellan verksamhetens effekter på dess omgivning å ena sidan, och Skatteverkets mål och visioner å andra sidan. I del III redogör vi, bland annat med exempel från vår verksamhet, för användbara kvantitativa respektive kvalitativa metoder. I referenslistan finns exempel på avsevärt mer uttömmande handböcker i metodik.

Eftersom rapporten spänner över olika perspektiv av effektmätning har den också olika adressater. Den som vill ha en praktisk handledning i att mäta effekter av en viss insats kan nöja sig med att efter sammanfattningen läsa kapitel 3, *Att fundera över effekter*, respektive hela del III, *Metoder för effektmätning*. Den som är beställare av en insats kan utöver sammanfattningen läsa kapitel 3 och 11. Del I och II behandlar effektmätning ur ett mer strategiskt perspektiv.

Vi hoppas kunna visa på ett överskådligt sätt vad man bör tänka på när man vill mäta effekter, vilka metoder som står till buds samt vilka fällor som man bör undvika. Detta omfattar många och ibland något teoretiska frågeställningar, som vi har försökt att konkretisera så långt varit möjligt. Redogörelsen är tänkt att fungera som ett hjälpmedel såväl vid den inledande planeringen av insatser som vid den senare utvärderingen av desamma. Den är också tänkt att vara ett underlag för en diskussion inom Skatteverket hur effektmätning bör användas.

2 Varför är det intressant att mäta effekter?

Styrningen av statlig förvaltning idag är starkt influerad av den modell för resultatstyrning av offentlig verksamhet som växte fram under 1980- och 90-talet ursprungligen från Storbritannien, och som kallas *New Public Management* eller kort NPM. Särdragen i denna modell kan sammanfattas enligt följande:²

- Man flyttar fokus från prestationer till effekter

² Pollitt, 2003, refererad i SOU 2007:75, s. 93

- Man betonar kvantifierade resultat, särskilt i form av prestationsmätningar
- Man inrättar beställar-/utförarorganisationer istället för traditionella hierarkiska organisationer
- Man ökar användningen av marknadsliknande mekanismer i olika former av serviceproduktion, ofta med betoning av kvalitet och brukarintressen

Som ett tydligt exempel på betoningen på brukarintressen fick vi för snart ett decennium sedan en förvaltningspolitisk proposition som betonade brukarupplevd kvalitet, och i Skatteverket har detta varit en ledstjärna ännu längre.³

Dessutom är myndigheter enligt 3 kap. 1 § förordningen (2000:605) om årsredovisning och budgetunderlag, ålagda att redovisa prestationer och effekter i förhållande till mål som regeringen ställt upp. Enligt den projektstyrningsmodell som används i hela Skatteverket sedan 2005, Pej1, ska det utföras effektanalyser och effektuppföljning av varje insats som utförs i projektform, låt vara att man i Pej1 även talar om ”interna” effekter, dvs. insatsers effekter på Skatteverkets egna verksamhet. Effektmålen enligt Pej1 ska utgå från de så kallade SMART-kriterierna, dvs. att målen ska vara Specifica, Mätbara, Acceperade, Realistiska och Tidsatta. I det årliga regleringsbrevet som innehåller regeringens styrdirektiv till Skatteverket, ges också ibland uppdrag att göra utvärderingar av effekterna av särskilda insatser, regel-förändringar etc.

I fortsättningen av detta kapitel ska vi först definiera några begrepp, för att därefter ge några exempel på teorier och metoder för effektmätning. Innan vi går vidare med en diskussion kring kopplingen mellan Skatteverkets strategier, mål och vision å ena sidan, och effektmätning å andra, ska vi först diskutera något om metoder för att mäta effekter, samt frågor om ansvarsfördelning. Detta kapitel avslutas med att vi lyfter blicken mot vad olika forskare kritiserar med effektmätning, generellt och som det sker i statlig verksamhet idag.

2.1 Effekter, resultat, prestationer – vad är det för skillnad?

En ofta citerad definition på effekter lyder som följer:

En effekt är en förändring som inträffat som en följd av en vidtagen åtgärd och som annars inte skulle ha inträffat (ESV 2006:8).

Ett vanligt begrepp i dessa sammanhang är det *kontrafaktiska tillståndet*, dvs. det tillstånd som skulle ha varit fallet om man inte hade vidtagit åtgärden. Det är således viktigt inte bara att tillståndet har förändrats – förändringen ska vara betingad av insatsen i sig. Som kommer att framgå är detta ofta ganska svårt att visa, och det kräver såväl rejäl eftertanke som resursförberedelse för att man ska lyckas.

Det är också viktigt att skilja på prestationer, dvs. vad verksamheten utför, och effekter, dvs. vad prestationerna leder till ute i samhället, för dess brukare etc. I engelskspråkiga rapporter brukar man skilja mellan å ena sidan *Output* (prestation), som ett mått på *Efficiency* (kostnadseffektivitet) och å andra sidan *Outcome* (effekt), som ett mått på *Effectiveness* (om insatsen har lett till måluppfyllelse).⁴ Det inre perspektivet (*efficiency*) mäter om man har gjort saker rätt, medan det yttre perspektivet (*effectiveness*) mäter om man har gjort rätt saker. Dessa två perspektiv är viktiga att hålla någorlunda åtskilda, bl.a. eftersom det behövs olika information för att bedöma om man har lyckats med uppgiften.

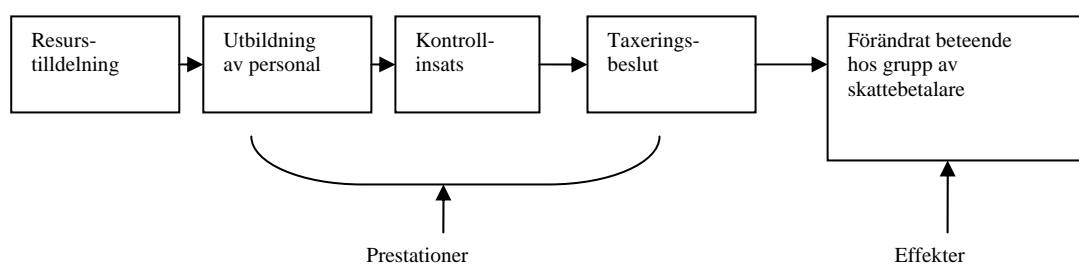
³ Prop. 1998/99:136, Statlig förvaltning i medborgarnas tjänst, RSV Rapport 1997:1, Resultatdialogen mellan RSV och SKM

⁴ Australian Taxation Office, 2007, s. 5.

Man skulle i och för sig kunna göra en uppdelning i ”interna effekter” (prestationer och deras utfall internt i organisationen vid olika insatser) och ”externa effekter”(vad som händer i omgivningen). Men eftersom intresset för effektmätning har ett nära samband med modeller för styrning av offentlig förvaltning, och begreppet ”effekt” i dessa modeller är avsett att specifikt fokusera intresset på vad som händer utanför organisationen, kommer vi fortsättningsvis i denna skrift att upprätthålla betydelskillnaden mellan begreppen prestation och effekt. Principerna för mätning av interna eller externa effekter är givetvis desamma, men de praktiska problemen med effektmätning uppstår i större utsträckning när det är effekter på Skatteverkets omgivning vi vill mäta.

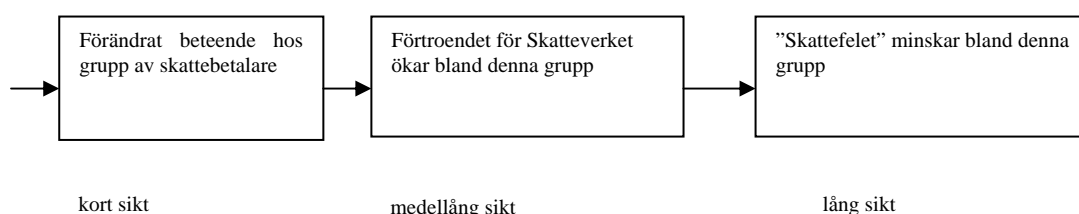
Skillnaden mellan prestationer och effekter kan åskådliggöras enligt följande enkla effektkedja, där innehållet varje enskild ruta förutsätts orsaka innehållet i nästkommande ruta till höger:⁵

Fig. 2.1 Enkel modell av effektkedja



I teorin brukar man dessutom skilja mellan kortsiktiga, medellånga och långsiktiga effekter. Effektkedjan ovan skulle exempelvis kunna delas in som följer:⁶

Fig. 2.2 Effekter på kort, medellång respektive lång sikt



I detta sammanhang ska det nämnas att det som en generell regel är svårare att konstatera ett samband ju längre ”avståndet” i orsakskedjan är mellan insatsen och den eventuella effekten. Detta beror helt enkelt på att det tillkommer fler och fler tänkbara faktorer som kan ha påverkat den tänkta effekten i någon riktning.

3 Att fundera över effekter

I utvärderingssammanhang brukar man betona vikten av att fundera kring vilken utvärderingsmodell och vilket perspektiv man väljer, eftersom detta kommer att styra över

⁵ Prestationsboxarna kan också delas in i 1) aktiviteter (utbildning och kontrollinsats) och 2) resultat (taxeringsbeslut).

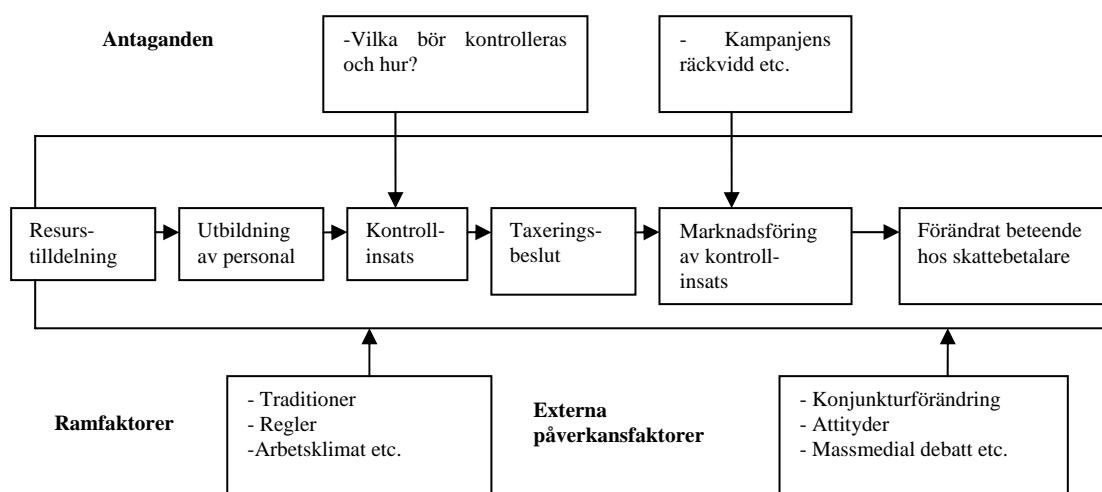
⁶ Notera att resonemanget ovan bara är ett exempel; man kan mycket väl argumentera för att det är ett ökat förtroende på kort sikt, som leder till ett ändrat beteende på medellång sikt, som på längre sikt resulterar i ett minskat skattefel för denna grupp.

vilka frågor som kan besvaras. Istället för att redogöra för dem alla, väljer vi att introducera en modell, här kallad utbyggd verksamhetslogik, som vi tror ofta passar vår verksamhets kunskapsbehov.

3.1 Verksamhetslogik

En bra utgångspunkt när man funderar kring hur man ska uppnå önskvärda effekter, alternativt vilka effekter som har uppnåtts och varför, är att skissa på en verksamhetslogik likt effektkedjan ovan i figur 1.1-1.2.⁷ Men denna bild behöver kompletteras med antaganden om varför effekterna uppstår, t.ex. hur utbildningen och kontrollen ska utformas, vilka som ska kontrolleras, hur insatsen ska marknadsföras etc. Dessutom bör bilden kompletteras med s.k. ramfaktorer, dvs. antaganden om vilka organisatoriska faktorer som kan tänkas påverka insatsens utformning och effekter, samt externa påverkansfaktorer, dvs. faktorer utanför organisationen som påverkar insatsens möjliga effekter. Det väsentliga är att man klargör så långt möjligt vilka bakomliggande antaganden som finns i ett projekt eller före en insats, samt att man försöker identifiera vilka faktorer som spelar in när det gäller att uppnå en önskad effekt. Om detta görs på planeringsstadiet, kommer man att formulera en verksamhetsteori som förtydligar de antaganden som finns i verksamheten om vad olika åtgärder leder till respektive vilka interna och externa påverkansfaktorer som kan identifieras. Det finns flera fördelar med detta arbetssätt än att dokumentera verksamhetens bakomliggande antaganden, bl.a. kan detta arbetssätt skapa en gemensam verksamhetsbild för dem som arbetar inom samma projekt.⁸

Fig. 3.1 Enkelt exempel på verksamhetslogik med ramfaktorer



Det är också viktigt att man redan på planeringsstadiet funderar kring vad och hur man avser att mäta effekterna, vilket också Pejl-modellen föreskriver för insatser som utförs som projekt. Om man underlåter detta och påbörjar insatsen ändå, kan det bli ännu svårare eller rentav omöjligt att i efterhand utföra mätningar på ett sätt som har förutsättningar att kunna säga något om effekter har uppnåtts eller inte.

3.2 När inträffar effekten, och hur länge varar den?

Samtidigt som man skissar på antaganden om hur en insats kan leda till en önskad effekt, är det värdefullt att man även försöker svara på när man kan förvänta sig att en effekt inträder,

⁷ I den svenska litteraturen kallas detta ibland även för programteori, jfr. Faugert & Sandberg, s. 88f.

⁸ En introduktion till att arbeta med verksamhetslogik finns utförligare beskrivet i ESV 2001:16.

och om möjligt, hur länge den kan antas vara. Detta får betydelse för när man planerar att göra en mätning (först när effekten kan förväntas inträda) respektive när man ser ett behov av att göra om en åtgärd (först när effekten av ursprungsinsatsen har klingat av).

3.3 Attributionsproblemet

När man funderar kring effekter, såväl i planeringsstadiet av en insats som i den senare utvärderingen, stöter man snart på det s.k. attributionsproblemet. Eftersom vi ytterst ska bedöma om en insats har påverkat ett beteende, behöver orsaksrelationen studeras. Det finns tre situationer då det är svårt att bekräfta om en observerad (eller förväntad) förändring kan härledas till en insats:

1) Förändringen skulle ändå ha inträffat på grund av t.ex. externa faktorer.

Man kanske kan mäta ett förändrat beteende, men samtidigt påtalas att förändrad lagstiftning i princip omöjliggjort det gamla beteendet. Insatsens effekt blir då kanske bara symbolisk.

2) Även om insatsen var nödvändig för förändringen av beteendet, var den kanske inte tillräcklig att ensam förändra detta. Detta är synnerligen viktigt vid de fall då man funderar på om ett resultat är överförbart till andra miljöer.

En insats mot en konjunkturkänslig bransch får effekt samtidigt som konjunkturen förändras. Det blir då svårt att veta om man kan förvänta sig samma effekt om insatsen istället utförs mot en bransch som är mer okänslig för konjunkturförändringar, eller i samma bransch under andra konjunkturrella förutsättningar.

3) Om man inte har kunnat mäta någon förändring i beteendet var kanske insatsen ändå viktig för att förhindra en *negativ* utveckling. Därför är det betydelsefullt att om möjligt analysera det kontrafaktiska tillståndet – hur det hade sett ut om insatsen inte hade vidtagits.

Detta kan inträffa t.ex. återigen vid konjunkturförändringar och kan kräva någon form av kvalitativ ansats, t.ex. intervjuer, för att man ska kunna bekräfta att en insats hade effekt trots att siffrorna kanske är oförändrade.

3.4 Evidensdiskussionen

För ett par decennier sedan debatterades inom forskarvärlden om huruvida samhällsvetenskapen skulle tillämpa kvantitativa eller kvalitativa metoder. Väldigt förenklat kan kvantitativa metoder sägas vara dem som använder matematiken (statistik etc.) som teknik för att analysera olika fenomen, medan kvalitativa metoder använder det talade och skrivna språket (intervjuer, observationer etc.). Det finns stora skillnader i vilken kunskapsframställning som är möjlig beroende på vilken av dessa metoder man använder. Utan att bli alltför teoretisk kan det sägas att kvantitativa metoder strävar efter objektivitet och därför göra åtskillnad mellan fakta och värde, mellan den studerande och det studerade. Laboratorieexperimentet är väl ett idealtypiskt exempel, där forskaren kan ha full kontroll på vilka faktorer som påverkar det studerade objektet. Kvalitativa metoder brukar utgå från det subjektiva och framhålla enhetligheten mellan fakta och värde, mellan den studerande och det studerade. Som ett exempel på det sistnämnda kan nämnas djupintervjuer, vilka ofta används inom vetenskapsområden som t.ex. etnografi. Intervjuerna används exempelvis för att beskriva intervjupersonernas kulturella uppfattningar om verkligheten – inte verklighetens utformning i sig.

Metoddebatten av idag har kretsar kring frågan om olika metoder kan rankas i en hierarki, där metoderna högt i hierarkin uttalar sig med större säkerhet, och de lägre metoderna med mindre säkerhet. De högt rankade metoderna anses bättre på att etablera det kontrafaktiska tillståndet. Evidensbaserad praktik inom samhällsvetenskaperna har vuxit fram på bred front, och syns bl.a. genom bildandet av olika myndigheter (t.ex. Institutet för metoder i socialt arbete) och nätverk (Campbell Collaboration) som med hjälp av s.k. metaanalyser (att analysera andra analyser) försöker svara på vad som fungerar och inte fungerar inom de stora välfärdsområdena socialt arbete, kriminalvård och skola.⁹ I skrifter om effektutvärdering är denna typ av hierarkier inte ovanliga, inte heller i skattemiljöer.¹⁰ Denna evidenshierarki brukar ranka metoder enligt figur 3.2 nedan.¹¹

Fig. 3.2 Exempel på evidenshierarki

1. Slumpmässiga kontrollerade experiment
2. Kvasi-experimentella ansatser
3. Före-efter-mätningar
4. Tvärsnittundersökningar
5. Processutvärderingar, aktionsforskning etc.
6. Kvalitativa fallstudier, etnografi etc.
7. Exempel på best-practises
8. Expertvärderingar
9. Brukarvärderingar

Konsekvensen av denna form av rankning är att de kvalitativa metoder som hamnar längre ned på listan, ofta döms ut som för osäkra i effektmätningssammanhang. De må vara mer osäkra för att isolera effekter av enskilda variabler, men har å andra sidan andra fördelar. Kvalitativa metoder kan bättre svara på frågor om varför en åtgärd ledde till någon viss effekt. När kvalitativa respektive kvantitativa metoder jämförs brukar man säga att de kvalitativa är teorigenererande, medan den kvantitativa är teoriprovande. Vi anser, med visst stöd, att man bedömer de olika metoderna utifrån dess egna premisser och använder dem till det som de är bäst för.¹² Under ideala omständigheter har man möjlighet att använda båda metodikerna. Vi kommer att redogöra mer för dessa i metodgenomgången i kapitel 3.

3.5 Ramar för effektmätning

3.5.1 Kostnad och tid

Som har antytts är det avsevärt svårare att mäta effekterna av Skatteverkets insatser, i jämförelse med att mäta prestationerna i form av antal revisioner, upptaxeringar, informationskampanjer etc. Svårigheterna består av att datatillgången normalt sett finns utanför verksamheten, och måste tas in separat. Därför är det normalt sett dyrare att få kunskap om effekter, och avsevärt mer tidskrävande än att mäta prestationer. Detta är viktiga faktorer att ta med i beräkningarna när man planerar för vad som ska effektmätas. Att göra effektmätning till en rutinuppgift kan, om dess fördelar inte blir synliga, leda till att man i organisationen tappar tilltro till ansatsen som helhet vilket vore en olycklig (och säkerligen kostsam) utveckling. Eftersom effektmätning är resurskrävande är det ofta inte kostnadseffektivt att försöka mäta effekter i för korta cykler.

⁹ Faugert & Sandberg, s. 75

¹⁰ Jfr. t.ex. ESV 2006:8, kap. 2, ATO, 2007, s. 46.

¹¹ Rieper & Foss Hansen, s. 20

¹² Denna ståndpunkt delar vi med till exempel Rieper & Foss Hansen, s. 87, och Världsbankens oberoende utvärderingsgrupp, IEG, s. 2.

3.5.2 Etik

Det är också viktigt att fundera kring om det kan finnas etiska hinder mot att anlägga en mer experimentell ansats inom ramen för en myndighets verksamhet. Man kan diskutera i vilken utsträckning mer experimentella ansatser främjar en god dialog med medborgarna.¹³

Medborgare i vårt samhälle har en lagstadgad rätt till likabehandling, vilket givetvis sätter gränser för vad man får göra med dem. Å andra sidan kan man även hävda att det är etiskt tveksamt att utföra åtgärder utan att veta om dessa gör någon skillnad. Användningen av personuppgifter för att göra slumpurval etc. för effektmätningar måste göras efter noggranna övervägningar. En generell råd är därför att rådgöra med en jurist om det kan finnas några juridiska hinder mot att utföra en effektmätning på planerat sätt. Som ett ytterligare exempel kan nämnas att alla myndigheter som vill utföra attitydundersökningar till företag är skyldiga att före undersökningen samråda med Näringslivets regelnämnd, NNR, enligt förordning (1982:668) om statliga myndigheters inhämtande av uppgifter från näringsidkare och kommuner.

3.6 Vem ska ansvara för effektmätning?

Det finns olika fördelar med olika placeringar av ansvaret för effektmätningar, och det bör finnas en medvetenhet om detta. Ansvaret för mätningarna kan antingen placeras hos samma grupp som har organiserat och/eller utfört den insats man vill mäta effekterna av, eller hos någon annan extern grupp. Vilket som är att föredra beror på vad man strategiskt finner vara de viktigaste fördelarna som man vill uppnå.

I utvärderingslitteraturen diskuteras ibland för- och nackdelarna med att utvärdering utförs inom en organisation eller av någon utomstående. Argumenten kan till viss del även appliceras inom organisationen, i relationen beställare – utförare, även om alla argumenten inte är applicerbara fullt ut. Vi avser med ”extern” någon som har en annan inomorganisatorisk tillhörighet än de som har utfört den insats som ska effektmätas.

Om utförarna själva ansvarar för att en effektmätning utförs, tar man fasta på den kunskap som finns om insatsen, det kan därför bli gjort relativt snabbt och kostnadseffektivt (givet att man har adekvat kompetens för effektmätningar). Nackdelarna är att det föreligger en risk för hemmablindhet och partiskhet, att interna konflikter påverkar mätarbetet samt att mätningen tar tid från andra uppgifter. Fördelarna med att låta någon utomstående från insatsen mäta dess effekter är att man tillför öppna ögon (motsatsen till hemmablindhet), man kan använda specifik mät- och metodkompetens samt att mätningen i bästa fall kan utföras utan ovidkommande intressen i resultatet. Ansvarsplaceringens för- och nackdelar kan delas in i följande matris:¹⁴

Fig 3.3 För- och nackdelar med intern respektive extern effektmätning

	Intern	Extern
Fördelar	Kunskap Snabbhet Kostnadseffektivt	Öppna ögon Specifik kompetens Opartiskhet
Nackdelar	Hemmablindhet Partiskhet Interna konflikter Resurskrävande	Dyrt Kontrollerande Brist på delaktighet

¹³ Faugert & Sandberg, s. 72.

¹⁴ Matrisen är modifierad från Faugert & Sandberg, s. 47

I Skatteverket är det beställaren av en åtgärd som har ansvaret för effektmätningen. Om man som beställare överväger hur effektmätning ska gå till i det enskilda fallet, finns det på huvudkontorets produktionsavdelning en arbetsgrupp som bistår med metodstöd, där bl.a. denna rapport författare ingår.

3.7 Fördelar med effektmätning

En av fördelarna med att utgå från mål och effekter är att verksamheten bättre kan identifiera vilka åtgärder som behöver vidtas för att nå dess övergripande mål, än om man utgick från åtgärder som aldrig ifrågasatts. På detta sätt kan man ta tillvara kompetens i olika delar av verksamheten och skapa delaktighet bakom verksamhetens mål.

En annan fördel med att arbeta systematiskt med effekter är att detta är en bra grund för ett lärande arbetssätt. I och med att verksamhetens effekter diskuteras och dokumenteras kan detta öka såväl kunskanden i som den gemensamma synen på verksamheten. Dessa fördelar ska inte underskattas.

3.8 Risker med effektmätning – vad säger kritikerna?

Kritikerna till trenden med effektmätning har framfört ett flertal argument mot att detta styrinstrument leder verksamheten mot något bättre. Detta har t.ex. diskuterats i nyligen publicerade rapporter från ESV, Verva och i den senaste offentliga utredningen om statlig styrning.¹⁵ Argumenten kan inledningsvis grupperas kring två huvudlinjer: för det första finns det olika svårigheter att med säkerhet belägga att en insats har haft effekt, och för det andra finns det empiriska belägg för att själva mätfokuseringen riskerar att styra verksamheter i oönskad riktning. Vi ska belysa några av kritikernas argument.

3.8.1 Osäkra mätningar

Ett problem med effektmått är dess osäkerhet, vilket återknyter till attributionsproblemet ovan: hur vet man att en given effekt är orsakad av insatsen? I utvärderingslitteraturen betonas ofta, för att inte säga alltid, att det är svårt att mäta effekter och en anledning till detta är att nödvändig data sällan finns omedelbart tillgänglig. Andra skatteförvaltningar har kommit till slutsatsen att det är svårt att mäta insatsers effekter på övergripande mål, jfr. t.ex. Australian Taxation Office:

Vi kan mäta kostnadseffektiviteten av våra insatser och nivån på allmänhetens förtroende för vårt arbetssätt, men vi har svårt att mäta hur effektiva våra strategier är i att förbättra skattebetalarnas beteenden. [...] Trots ett flertal individuella och samfälliga studier utförda av internationella skatteverk, går det långsamt (vår översättning).¹⁶

Eftersom data inte sällan är osäker, och mätningar ställer tämligen höga krav på metodkunskap kan det bli dyrt att mäta effekter. Vi har redan berört bland annat kostnaden som en av ramarna för effektmätningar, och kostnaden måste ställas i relation till den förväntade möjligheten att få svar på frågan om en insats har haft effekt och i så fall vilken.

För komplexa verksamheter som offentliga tjänster är effektfrågan utomordentligt svårlöst. Det är därför inte en slump att effektfrågan är ett kärnproblem i utvärderingssammanhang

¹⁵ ESV 2007:23, Verva 2007:12, SOU 2007:75

¹⁶ "We can measure the efficiency of our compliance strategies and the level of community confidence in our approaches but we have difficulty measuring how effective our strategies are in improving taxpayers' compliance behaviour. [...] Despite a number of individual and cooperative studies conducted by international agencies, progress is slow." ATO, s. 5

som diskuteras i alla böcker i ämnet. [...] Empiriska studier visar att det är sällsynt att effektfrågan alls problematiseras när [kvalitets- och resultatmätningssystem] konstrueras.¹⁷

Och även om data finns tillgänglig måste man vara uppmärksam på huruvida man kan isolera effekten av den variabel (den företagna åtgärden) som man vill mäta. Hur detta kan gå till rent praktiskt diskuteras mer i del III av denna rapport.

Som beskrivits ovan under ramar för effektmätning, måste man ställa kostnaderna för effektmätning i relation till förväntade utsikter att kunna uttala sig om effekterna med någon säkerhet. Om man försöker att tumma på kvalitetskraven i mätningarna för att minska kostnaderna för dessa, riskerar man att resultaten ifrågasätts och att tilltron till effektmätning som verktyg sjunker, vilket vore en olycklig utveckling. Därför bedömer vi att man måste pröva från fall till fall om de planerade insatserna ska effektmätas. Dessutom kan kravet på mätbarhet leda till oönskade bieffekter.

3.8.2 Målförskjutning

Kritiker mot ökad användning av uppföljning och utvärdering som styrmedel diskuterar ibland faran med oönskade bieffekter av mätning, alltså mätning av såväl prestationer som effekter. Att effektmål ska vara SMART:a har kritiserats från en mängd håll. Det är främst mätbarhetskravet som styrmedel som kritiserats.¹⁸ Som nämnts ovan framhålls målstyrning och effektmål som ett styrsätt som främjar kreativitet, eftersom utförare får flexibla möjligheter att själva utforma aktiviteter efter vad de anser bäst skapar möjlighet att nå målen.¹⁹ Här kan emellertid mätbarhetskravet ha en hämmande effekt. Kritiken formuleras av vissa som att det kan uppstå en konflikt mellan det mätbara och det väsentliga, och då vore det olyckligt om det mätbara ”vann”. Detta innebär en form av målförskjutning. Så här uttrycker en ledande företrädare för utvärderarkåren det:

När indikatorer blir mål i sig resulterar detta i målförskjutning, som leder till att tonvikten läggs på fel aktiviteter och uppmuntrar till ”creaming” [att mäta det som på förhand antas ge bra resultat] och andra sätt att ”klara siffrorna” utan att egentligen förbättra faktiska effekter av en verksamhet (vår översättning).²⁰

I massmediadebatten förekommer när detta skrivs diskussioner om hur lärare och rektorer känner en stark press att höja elevens betyg, i syfte att upprätthålla nöjda föräldrar och att stärka skolans finanser. Detta beteende, som är ett slags målförskjutning, bidrar emellertid inte till att förbättra elevernas kunskaper eller förmåga att fungera som samhällsmedborgare.

Om mål är vagt formulerade och det samtidigt finns ett krav på mätbarhet, riskerar man att planeringsprocessen handlar om att först inventera vad som kan mätas, och sedan utföra respektive mäta detta. Detta kan sägas vara motsatsen till ursprungstanken bakom att styra mot effekter, som förevisar att man ska utgå från önskade effekter och sedan utföra det man tror leder mot dessa.²¹ Ett klassiskt men kanske drastiskt exempel på ett missvisande och målförskjutande mått härstammar från Vietnamkriget, där den amerikanska regeringen använde ”antal dödade fiender” som en indikator på framgång. Detta styrmått ansågs sedermera vara en delorsak till tragiska händelser där militära ledare som ansåg sig skyldiga

¹⁷ Lindgren, s. 90

¹⁸ ESV 2006:7, s. 12, Verva 2007:12, s.149, Tarschys, refererad i SOU 2007:75, s. 162.

¹⁹ Jfr t.ex. Pihlgren & Svensson, s. 10

²⁰ ”When indicators become the objective, they result in ‘goal displacement’, which leads to emphasis on the wrong activities and encourage creaming and other means of ‘making the numbers’ without improving actual outcomes.”, Perrin, s. 371

²¹ Lindgren, s. 87

att leva upp till siffrorna orsakade massakrer på orsbefolkningar.²² Detta beteende ledde till stora framgångar enligt mätningen, men inte till att man vann kriget.

3.8.3 Andra exempel på oönskade bieffekter – avprofessionalisering och känsla av övervakning

Det talas inom forskningen även om andra oönskade bieffekter av verksamhetsstyrning. Redogörelsen nedan är inte någon fullständig förteckning, utan endast ett sätt att försöka visa på att det behövs en medvetenhet om hur organisationer fungerar när man ska försöka förklara om och varför ett visst resultat har uppnåtts (eller inte). Därför bör detta nämnas i ett effektmätningssdokument.

En önskad bieffekt med att styra mot det mätbara, är risken för avprofessionalisering.²³ Resonemanget bygger på situationen där mätningen griper in i en professionell sfär, där de professionella anser sig äga kriterierna på vad som är bra eller dålig verksamhet. Genom införandet av mått och indikatorer som fastställs av externa instanser, finns en risk för att professionen reagerar med misstro och börjar tvivla på sin professionalism. Om avprofessionalisering inträffar, finns en risk att en insats önskade effekter såsom t.ex. bättre brukarbetyg inte uppnås. Detta borde i och för sig inte vara ett oöverstigligt hinder, men det sätter fokus på vikten av att förankra mätmetoder och –rutiner hos dem vars arbete ska bedömas.

En annan bieffekt av ökad resultatmätning generellt är risken att de utvärderade (personalen) i större utsträckning känner sig övervakade, vilket i värsta fall kan leda till att man inte vågar ifrågasätta vad man anser vara brister med verksamheten. Ytterst innebär detta att verksamheten riskerar att förstelnas, genom att kunskapen om nackdelar med verksamheten inte kommer ledningen till godo. Detta är en kontraproduktiv bieffekt som motverkar en av fördelarna med att styra mot effektmål, att uppmuntra till kreativa lösningar på verksamhetens problem.

3.9 Konflikt mellan det väsentliga och det mätbara?

Som beskrivits ovan framhålls i flera nyligen publicerade rapporter, och bland forskare sedan åtminstone något decennium att fördelarna med att styra mot mål och effekter - tillvaratagandet av kompetens, skapande av kreativitet till gagn för verksamhetens målgrupper, bättre lärande inom organisationen – riskerar att gå förlorade på grund av bieffekterna som kan uppstå om man upprätthåller ett krav på att effekter alltid ska vara mätbara för att åtgärden ska kunna vidtas. Detta torde vara en särskilt stor risk när det som ska mätas, i vårt fall effekter, är svårt på grund av datatillgång, komplexitet etc.

Vi menar inte att man med anledning av detta ska undvika att styra mot effektmål på grund av att de är svåra att mäta. Tvärtom hävdar vi att effektmål är bra. Det är däremot inte bra när mätning av prestationer får ersätta effektmålen som styrmedel. Produktivitet och kvalitet kan behöva mätas av andra anledningar.

Vi rekommenderar att man inom Skatteverket överväger nyttan med kravet att effekter av alla åtgärder ska vara mätbara för att man åtgärderna ska få utföras. Däremot är det lämpligt att man inför varje åtgärd utreder om effekterna går att mäta. På detta sätt är det möjligt att allokeras mätresurser till strategiskt viktiga projekt och insatser för att mer kostnadseffektivt

²² Campbell, s. 54

²³ Lindgren, s. 108 f.

vinna kunskap. Som verksamhetsansvarig bör man emellertid alltid föra en dialog om varför man anser att en åtgärds effekter inte är mätbara/bör mätas.

Del II: Effektmätning utifrån Skatteverkets övergripande mål

Skatteverkets yttersta uppgift är att se till att den demokratiskt beslutade skatten fastställs och betalas på avsett sätt. I och med att det är den yttersta uppgiften är det också målet och meningen med alla de aktiviteter som företas inom verket. Alla aktiviteter ska alltså sist och slutligen på något sätt bidra till att det övergripande målet nås.

Syftet med detta avsnitt är att undersöka möjligheterna att mäta de enskilda aktiviteternas effekter på det övergripande målet. Det är viktigt att hela tiden ha i minnet att resonemanget enbart utgår från möjligheterna att mäta olika insatserns effekter. Insatser och mål betraktas i detta sammanhang som givna. Att *insatserna* betraktas som givna innebär att det inte förs något resonemang om insatserna i sig, vilken mängd som är optimal eller hur de ska utföras. Med andra ord så behandlas överhuvudtaget inte frågor om *produktivitet* eller *kvalitet* i detta dokument. Att *målet* är givet betyder att utgångspunkten är det övergripande målet så som det formuleras av Skatteverkets uppdragsgivare regeringen, som i sin tur agerar på uppdrag av medborgarna, de yttersta uppdragsgivarna.

Detta avsnitt börjar med att undersöka hur Skatteverket sett på effektmätning i de verksamhetsuppföljningsmodeller som används och använts i verksamheten (kapitel 4). Därefter undersöks hur det övergripande målet kan betraktas ur effektmätningssynpunkt (kapitel 5). Den sista delen ger sig i kast med kopplingen mellan enskilda aktiviteter och det övergripande målet (kapitel 6).

4 Skatteverkets resultatuppföljningsmodeller och en idealisk effektmätningssynpunkt

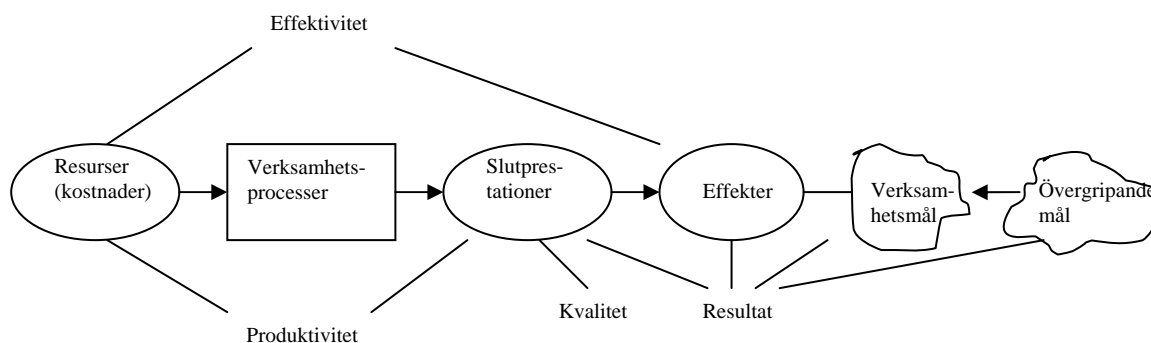
Skatteverket har alltid på ett eller annat sätt följt upp verksamheten och bedömt dess resultat. Vid olika tillfällen har det tagits fram modeller och verktyg för ändamålet. De två senaste är Resumé, som presenterades i en rapport 1997, och Guppi, som presenterades i en rapport år 2004.²⁴ Resultatuppföljningsmodeller är något annat än effektmätning. De avser att mäta resultatet av hela verksamheten, dvs av alla samlade insatser som görs. De beaktar också de två viktiga aspekterna *produktivitet* och *kvalitet*. Effektmätning är i och med sin avgränsning till enskilda insatser som är givna något mycket mer begränsat än resultatuppföljning. Hur Skatteverket har sett på respektive ser på resultatuppföljning är ändå i högsta grad intressant och relevant ur effektmätningssynpunkt.

4.1 Resumés bild av verksamheten och uppföljningen av den

Nedan visas den kedja från insatta resurser till effekter som användes som utgångspunkt när Resumés resultatmått togs fram.

²⁴ RSV Rapport, Resultatdialogen mellan RSV och SKM (Resumé), SKV Rapport 2004:9, Grundläggande uppföljningsinformation (Guppi)

Fig. 4.1 Grundläggande effektkedja i Resumémodellen



Resumé utgår från det övergripande mål som då fanns, ”rätt skatt på rätt sätt”. Detta mål bröts sedan ner till resultakriterier på tre nivåer. Den högsta nivån handlade om övergripande effekter, som inte var direkt mätbara. För beskattningsverksamheten var det högsta resultatkriteriet *beskattningseffekt*, dvs att minimera gapet mellan fastställd och teoretiskt riktig skatt. Under detta kom ett antal underkriterier på nivå två. Ett sådant kunde för beskattningseffekten vara *preventiv effekt*, vilken konstaterades inte heller vara mätbar. Därunder kom sedan resultakriterierna på nivå tre. Ett exempel på ett sådant för resultatkriteriet preventiv effekt var *risk för upptäckt*. Eftersom denna risk inte heller är direkt mätbar togs det fram ett antal resultatindikatorer. Indikatorer för risk för upptäckt kunde vara ändringsfrekvenser vid olika former av kontroller.

På detta sätt bröts det övergripande målet ”Rätt skatt på rätt sätt” ned på resultatindikatorer för alla olika verksamhetsgrenar. Till detta lades även produktivitet och kvalitet. En av Resumé-rapportens stora förtjänster är att den i detalj benär upp och på ett stringent och logiskt sätt visar sambanden mellan verksamhetens olika insatser och de övergripande målen. Till detta presenterades också ett sätt att mäta verksamhetens resultat genom indikatorer. Ett problem med Resumémodellen är att indikatorerna ofta utgörs av kontroll- och höjningsfrekvenser vilka, om de ensamma får utgöra styrmått för verksamheten, lätt kan leda till ett oönskat statistikbeteende och det slags målförskjutning som beskrivs i avsnitt 3.8.2. Detta är något som också påpekas i Resumérapporten upprepade gånger. Dock ansågs modellen kunna fungera som styrinstrument under förutsättning att det är den information som sammantaget fås av samtliga indikatorer som används för analys och styrning av verksamheten.

4.2 Guppis bild av verksamheten och uppföljningen av den

I de direktiv som gavs till den arbetsgrupp som hade till uppgift att ta fram en ny uppföljningsmodell, som senare resulterade i Guppi, angavs att effekter skulle vara i fokus. Arbetsgruppens arbete byggde vidare på tankarna i Resumé. En viktig strävan var dock att finna metoder att mäta verkliga effekter istället för prestationer.

Arbetsgruppen konstaterade att det är svårt men inte omöjligt att mäta effekter. De kan inte mätas med hög precision eller detaljeringsgrad och de kan endast mätas långsiktigt. Arbetsgruppen tog fram ett förslag på en s.k. basplatta som avser att mäta delar av skattefelet samt deffekter på vägen mot de övergripande målen. De eftersträvar effekterna beskrivs genom nedanstående bild.

Fig. 4.2 Guppimodellen - från åtgärder till övergripande mål



Det som föreslås ska mätas i basplattan är bl.a. skattefelet inom vissa områden samt normer och attityder. Skattefelet är tänkt att mätas genom t.ex. slumpmässiga urval och nollnivåer i urvalssystemen. Normer och attityder mäts genom enkätfrågor.

Av bilden ovan kan utläsas att arbetsgruppen betraktar sambandet mellan Skatteverkets aktiviteter och prestationer å ena sidan och de övergripande målen å den andra som komplex, med ett nät av länkar och kopplingar mellan olika aktiviteter, prestationer och deffekter. En central del i modellen är att all strävan syftar till att skattebetalarna ska medverka frivilligt genom att göra rätt från början, något som innebär en restriktion för målen och medlen. Ett nyckelord i modellen är förtroende.

En viktig poäng med restriktionen ”rätt från början” är att Skatteverkets olika prestationer och aktiviteter inriktas och koncentreras på ett sätt som minskar risken för det slags målförskjutning som beskrivs ovan i samband med Resumé-modellen.

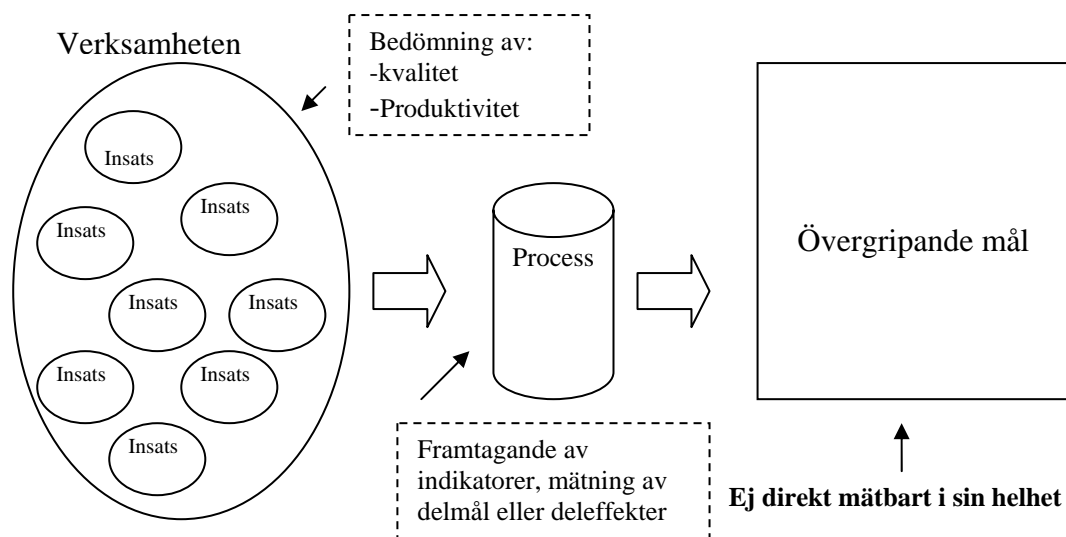
4.3 De viktigaste skillnaderna mellan Resumé och Guppi

I både Resumé-rapporten och Guppi-rapporten konstateras att det inte är möjligt att mäta de eftersträvade effekterna direkt. Resumés lösning på detta problem är att mäta prestationer som logiskt och erfarenhetsmässigt anses vara sammanlänkade med de övergripande målen. Guppi strävar ett steg längre och pekar på sätt att mäta om inte de eftersträvade sluteffekterna, så ändå olika deffekter som anses vara en del av de övergripande målen. En viktig skillnad mellan de två modellerna är att de direkta kontrolleffekterna har en viktig plats i Resumé medan kontrollen i Guppi endast utgör en indirekt påverkansfaktor. De indirekta effekterna av kontrollen i Guppimodellen beräknas uppstå genom flera olika mekanismer. En sådan mekanism är att fler anses göra rätt från början om den upplevda risken för upptäckt ökar. En annan mekanism är att förtroendet för Skatteverket och skattesystemet anses öka om kontrollen är effektiv. Det ökade förtroendet gör i sin tur att drivkraften hos skattebetalarna att göra rätt från början ökar.

4.4 Sammanfattning och slutsatser kapitel 4

De två verksamhetsuppföljningsmodellerna har alltså ett vidare och större syfte än en effektmättningsmodell. Verksamhetsuppföljningsmodellerna strävar efter att mäta effekterna av Skatteverkets alla aktiviteter och prestationer inom alla verksamhetsområden. I modellerna beaktas dessutom andra viktiga faktorer, såsom produktivitet och kvalitet. En effektmättningsmodell strävar efter att visa effekten av en enstaka prestation som är given. Att den är given betyder att sådana faktorer som produktivitet och kvalitet inte beaktas, inte för att det skulle vara oviktigt utan för att de ligger utanför modellen och förhoppningsvis beaktas på annat sätt. Skillnaden kan illustreras som nedan.

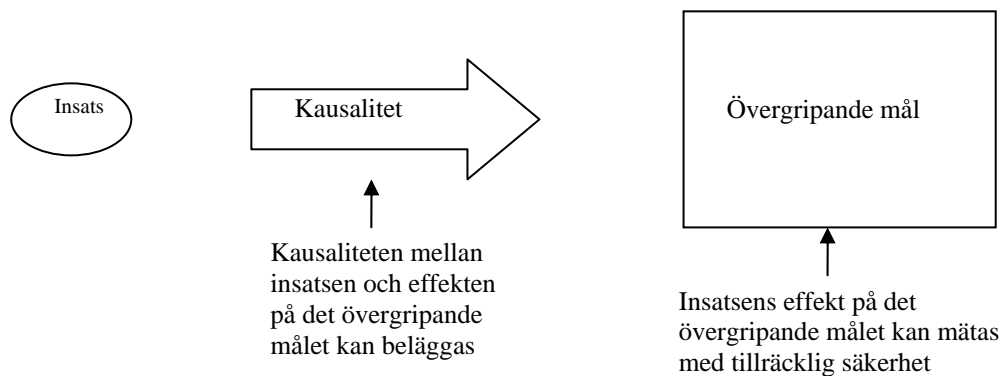
Fig 4.3 Insater och effekter enligt verksamhetsuppföljningsmodellerna



Både Resumé- och Guppi-modellen strävar efter att mäta effekterna av Skatteverkets verksamhet med avseende på de övergripande målen. I båda modellerna konstateras att effekterna på de övergripande målen inte är direkt mätbara. Guppi ser dock vissa möjligheter att mäta deffekter.

En idealisk effektmättningsmodell kan beskrivas som nedan. En insats leder till en effekt, som är mätbar, och sambandet mellan insatsen och effekten kan fastställas.

Fig 4.4 Insater och effekter enligt en idealisk effektmättningsmodell



Tyvärr är det lika omöjligt att mäta enskilda insatsers effekter på de övergripande målen som det är att mäta hela verksamhetens inverkan på dem. Utmaningen att ta fram en effektmättningsmodell kan dessutom anses vara större eftersom det finns ett krav på att det ska finnas ett kausalt samband mellan insatsen och effekten, dvs det är insatsen och bara insatsen som ska ha lett till effekten.

Därmed står den som har till uppgift att ta fram en effektmättningsmodell inför liknande, och inte mindre svåra, utmaningar som den som ska ta fram en verksamhetsuppföljningsmodell. Eftersom grunderna för verksamhetsuppföljning och effektmätning är desamma tar de fortsatta försöken att skissa på en effektmättningsmodell sin utgångspunkt från de båda verksamhetsuppföljningsmodellerna. Problemet med kausalitet behandlas vidare i avsnitt 6.1.

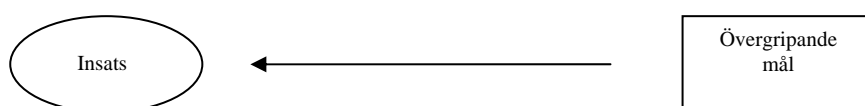
Eftersom det inte går att mäta sambandet mellan insatsen och de övergripande målen behandlas de två delarna, insatser och övergripande målen, separat. Först behandlas i kapitel 5

vad som faktiskt kan mätas av de övergripande målen utan krav på samband med enskilda insatser. Nästa del, kapitel 6, handlar om vad som kan mätas utifrån specifika insatser, utan krav på att måtten uttrycks i samma termer som de övergripande målen.

5 Strategiskt perspektiv - vad kan mätas med utgångspunkt i våra övergripande mål?

Denna fråga har redan utförligt behandlats och utretts av de båda arbetsgrupperna som tagit fram Resumé och Guppi. Båda har konstaterat att de övergripande målen i bästa fall kan följas på lång sikt. Det behöver därför inte spillas ytterligare kraft och energi på ett konstaterande att det inte går att mäta enskilda insatsers inverkan på de långsiktiga målen. Det bästa som kan göras är att med lämpliga tidsintervall försöka uppskatta måluppfyllelsen av de långsiktiga målen, frikopplat från eventuell påverkan från enskilda insatser. Detta görs också regelbundet i t.ex. årsredovisningen.

Vad som ska göras här är att föra ett mycket löst resonemang om hur långt det faktiskt går att bryta ner det övergripande målet i delar och vilken påverkan olika typer av insatser kan ha på dessa delar. Målet är att dra upp en skiss över en modell som skulle kunna vara användbar för effektmätningar och analyser på strategisk nivå.



Utgångspunkten är således de långsiktiga målen, och perspektivet är därmed i allt väsentligt strategiskt. Resonemanget strävar sedan bakåt mot olika insatsers eventuella effekt. I kapitel 6 görs det motsatta, dvs utgångspunkten är olika insatser och i vilken grad dessa kan anses påverka de långsiktiga målen. Detta kapitel har ett mer operativt perspektiv.

5.1 Övergripande mål

Effektmätningen ska inta ett externt perspektiv, dvs den ska inte handla om att mäta verksamheten i sig, dess delar eller interna processer. Effektmätningen ska istället utgå från det övergripande uppdraget. Utgångspunkten ska vara de yttersta uppdragsgivarnas perspektiv. Liksom för alla offentligfinansierade verksamheter är Skatteverkets yttersta uppdragsgivare medborgarna. Medborgarna ger i sin tur regeringen i uppdrag att styra myndigheterna. Målen för respektive myndighet anges i regleringsbrevet.

5.1.1 Formulering av övergripande effektmål utifrån regleringsbrevet

För politikområdet skatt, tull och exekution är målet i regleringsbrevet att: ”säkerställa finansieringen av den offentliga sektorn och bidra till ett väl fungerande samhälle för medborgare och näringsliv samt motverka brottslighet”.

För verksamhetsområde skatt anges målet: ”eftersträvade skatte- och avgiftsintäkter skall säkerställas rättssäkert och på ett för samhället ekonomiskt effektivt sätt samt på ett sätt som är kostnadseffektivt och enkelt för medborgare och företag. Brottslighet inom skatteområdet skall förebyggas och bekämpas. Medborgare och företag ska ha förtroende för Skatteverkets verksamhet.”

För verksamhetsgrenen beskattning anges: ”skatter och avgifter skall fastställas så att skillnaden mellan de fastställda och teoretiskt riktiga beloppen (skattefelet) blir så litet som möjligt. Fastställda belopp skall bli betalda i sin helhet och inom utsatt tid. Preliminärskatt skall i så liten utsträckning som möjligt avvika från den slutliga skatten (uppbördsfelet).”

De tre målen i regleringsbrevet innehåller mycket. För att i någon mån göra dem hanterbara ur effektmätningssynpunkt görs här en avgränsning. Den första avgränsningen är till verksamhetsgrenen skatt. Detta innebär att tull- och exekutionsväsendena inte ingår. För verksamhetsområdet skatt bortses även från fastighetstaxering, folkbokföring och brottsbekämpning. Vidare bortses från kraven på kostnadseffektivitet och rättssäkerhet. De senare är, som redan nämnts, sådant som definitionsmässigt faller utanför effektmätningssynpunkten.

Det som återstår är således att Skatteverket ska säkerställa finansieringen av den offentliga sektorn. Detta ska göras på ett för samhället effektivt sätt som är kostnadseffektivt och enkelt för medborgare och företag.

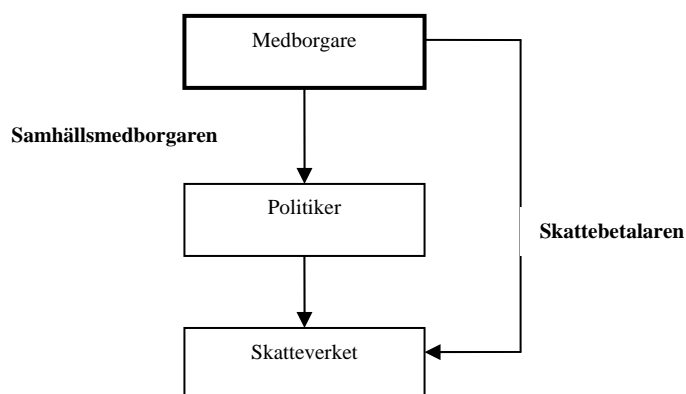
5.1.2 Medborgarens uppdrag till Skatteverket - rättighetssidan

Ur denna koncentration och avgränsning av regleringsbrevets mål framträder två olika uppdrag från medborgarna.

1. För det första ska medborgarna kunna lita på att de gemensamma nyttigheterna finansieras på avsett sätt.
2. För det andra ska varje enskild skattebetalare kunna utföra sina skyldigheter på ett så enkelt, billigt och smidigt sätt som möjligt.

Varje medborgare kan alltså sägas ge två uppdrag till Skatteverket, ett direkt och ett indirekt. Det direkta uppdraget är som skattebetalare. I denna egenskap har medborgaren en direkt kontakt med Skatteverket. Dennes förväntningar på Skatteverket är att kunna fullgöra sina skyldigheter med minsta möjliga insats. Det indirekta uppdraget är som samhällsmedborgare. I den egenskapen finns ingen direkt beröring med Skatteverket. Uppdraget ges genom det politiska systemet. Denna del handlar för medborgaren om att det ska finnas en tillit och ett förtroende för att Skatteverket utför sitt uppdrag så att de gemensamma nyttigheterna blir finansierade på avsett sätt. Ur den enskildes perspektiv handlar det kort sagt om en tilltro till att Skatteverket tar in den rätta och riktiga skatten för alla andra.

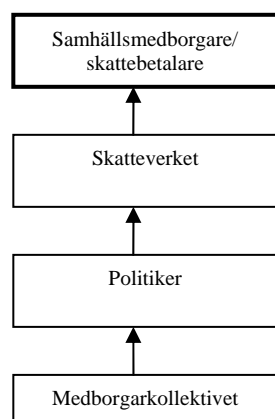
Fig 5.1 Medborgarnas uppdrag till Skatteverket - rättighetssidan



5.1.3 Medborgarens skyldigheter

Resonemanget ovan är begränsat till den enskilde medborgarens rättigheter. Implicit i regleringsbrevets uppdrag till Skatteverket finns dock även ett krav på den enskilde medborgaren att lojalt bidra med sin del i finansieringen av de gemensamma nyttigheterna. På samma sätt som den enskilde kan förvänta sig att alla andra medborgare ska betala sina skatter ska denne själv fullgöra sina skyldigheter i detta avseende. För den enskilde medborgaren finns alltså även en skyldighetssida. Skyldigheterna är gentemot medborgarkollektivet med Skatteverket och politikerna som mellanliggande led.

Fig 5.2 Medborgarens skyldigheter



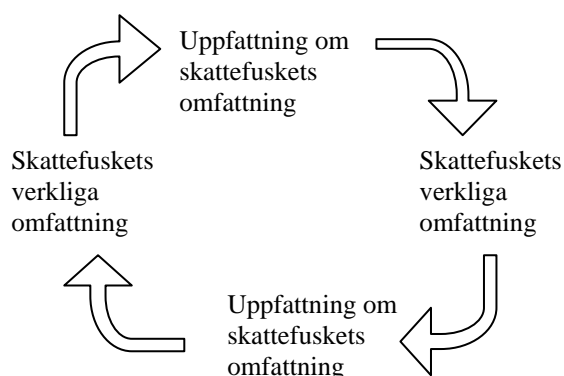
5.1.4 Måluppfyllelsen kan mätas med både subjektiva och objektiva mått

En vanlig metod när någon företeelse ska mätas eller undersökas är att fråga ett urval personer om deras uppfattning om saken. Så länge frågan handlar om personerna själva sammanfaller personernas uppfattning definitionsmässigt med det som ska undersökas. Så är fallet för t.ex. partisympatiundersökningar och andra åsiktsfrågor. I många fall går dock frågorna utöver personerna själva. Exempelvis ställs ibland frågor till allmänheten om deras uppfattning om tillståndet i samhället i stort, såsom kriminalitetens utbredning, hur det står till med jämlikheten på olika områden eller skattefuskets omfattning. Eftersom uppfattningen om olika företeelser mycket sällan överensstämmer med hur det faktiskt ser ut är det ur mätsynpunkt

ofta motiverat att skilja på upplevelsen av världen och världen i sig. Det handlar alltså om subjektiva och objektiva mått.

Båda måtten är ofta intressanta. Inte minst för att de kan samspela på ett svårförutsägbart sätt. Till exempel finns mycket som talar för att föreställningen om skattefusket har betydelse för skattefusket verkliga omfattning. Samtidigt kan det finnas en tröghet när det gäller hur förändringar i skattefusket verkliga omfattning slår igenom i allmänhetens medvetande.

Fig 5.3 Samband mellan föreställningen om skattefusk och skattefusket verkliga omfattning



Det som är huvudämnet för detta avsnitt, rättigheter och skyldigheter i ett skattebetalar- och samhällsmedborgarperspektiv, kan på samma sätt som ovan betraktas på två sätt. När det gäller medborgarnas rättigheter handlar det dels om medborgarnas subjektiva uppfattning om Skatteverkets förmåga att utföra sitt uppdrag, dels om olika objektiva mått på samma sak. Exempel på det senare kan vara att faktiskt försöka uppskatta i kronor, tid eller annan mätbar enhet i vilken grad Skatteverket lyckas med sin uppgift.

Hur Skatteverket utför sitt uppdrag är ett exempel på en fråga där det kan finnas en stor skillnad mellan å ena sidan den subjektiva uppfattningen och olika objektiva mått å den andra. Detta är naturligt eftersom saken inte kan överblickas i sin helhet av en enskild person.

Även när det gäller skattebetalarnas skyldigheter, som helt berör den enskilda personen, kan det finnas en poäng med att skilja på subjektiva och objektiva mått. Det kan nämligen finnas en skillnad i skattebetalarnas vilja att göra rätt och skattebetalarnas förmåga att göra rätt.

Alla dimensioner av det övergripande målet har enligt resonemanget ovan delats upp i en subjektiv och en objektiv sida. Denna uppdelning har bedömts vara värdefull ur informations- och analysynpunkt.

5.1.5 Det övergripande målet uppdelat i samtliga dimensioner

I det övergripande målet finns således både en rättighets- och en skyldighetssida.

På rättighetssidan har den enskilde medborgaren två intressen. Det första är att på ett så enkelt och smidigt sätt som möjligt kunna fullgöra sina skyldigheter. I detta avseende har den enskilde en direkt kontakt med Skatteverket. Den enskilde kallas i denna egenskap för skattebetalaren. Det andra intresset som den enskilde medborgaren har är att Skatteverket sköter sin uppgift att ta in den skatt som är demokratiskt beslutad. Detta uppdrag ges till Skatteverket av den enskilde genom det politiska och demokratiska systemet. Den enskilde kallas här för samhällsmedborgaren.

Skyldighetssidan handlar om den enskildes plikt att lojalt bidra till finansieringen av den skattefinansierade sektorn. Dessa olika dimensioner av det övergripande målet kan sedan mätas både genom subjektiva och objektiva mått. Samtliga dessa dimensioner av det övergripande målet visas nedan i matrisform.

Fig 5.4 Det övergripande målets olika dimensioner

Egenskap av	Rättigheter		Skyldigheter	
	Subjektiv bedömningsgrund	Objektiv bedömningsgrund	Subjektiv bedömningsgrund	Objektiv Bedömningsgrund
Skattebetalare	Mötet Enkätfrågor om t ex bemötande och fungerande system	Fullgörandet Mätning av fullgörandekostnader, diagnostiskt formulär för förenklingar, andel e-deklarationer, andel som överhuvudtaget haft kontakt med SKV etc.	Lojaliteten Enkätfrågor om viljan att göra rätt etc	Efterlevnaden Förekomsten av fel (avsiktliga, oavsiktliga och ignoransfel)
Samhällsmedborgare	Tilliten Enkätfrågor om förtroendet för skattesystemet, SKV etc.	Uppfyllelsen Skattefelskarta och medverkandekarta		

Varje kombination av medborgaregenskap, medborgarkrav och bedömningsgrund är i någon mening en egen aspekt eller dimension av det övergripande målet. I matrisen har varje sådan måldimension fått en egen benämning. Dessa är för medborgarens rättighetssida *Mötet*, *Fullgörandet*, *Tilliten* och *Uppfyllelsen*. För varje måldimension finns angivet ett antal mätvariabler och mål som kan användas för bedömning av dem. De flesta av dessa mäts redan regelbundet, och är således inget nytt. Det enda som möjligen är nytt är strukturen och något av de objektiva måtten. Skattebetalarens/samhällsmedborgarens skyldighetssida har det subjektiva bedömningskriteriet *Lojaliteten* och det objektiva *Efterlevnaden*.

Idén med en uppdelning av det övergripande målet på detta sätt är att finna mått som å ena sidan är tillräckligt strukturerade och koncentrerade för att vara hanterbara ur effektmätningssynpunkt, och samtidigt är tillräckligt dynamiska och mångfasetterade för att fånga in Skatteverkets komplexa verklighet i alla sina aspekter å den andra sidan.

Nedan beskrivs närmare de olika måldimensionerna och vilka mått som kan tänkas fånga in dem.

Skattebetalarens rättigheter

Skattebetalarens rättigheter handlar alltså om dennes rätt att kunna fullgöra sina åligganden som skattebetalare på ett så enkelt och effektivt sätt som möjligt. Den subjektiva sidan av detta kallas mötet, medan den objektiva kallas fullgörandet.

Mötet

Den måldimension som benämns *Mötet* handlar om hur den enskilde skattebetalaren upplever möjligheterna att fullgöra sina skyldigheter. Detta är något som redan mäts regelbundet, främst genom enkäter. Denna del fångas upp av frågor om bemötande, om det är lätt att göra rätt etc.

Fullgörandet

Fullgörandet är den objektiva sidan, som mäter vad som krävs av skattebetalarna för att de ska kunna fullgöra sina skyldigheter.

Ett objektivt mått på fullgörandet eller belastningen för den enskilde är fullgörandekostnaderna. Skatteverket genomför inte självständigt några regelbundna mätningar av fullgörandekostnaderna. Den senaste som gjordes handlade om kostnaderna för de momsskattelikliga företagen som gjordes 2005. Dessförinnan mättes 1999 privatpersonernas kostnader för att deklarerat. Däremot genomför Nutek löpande uppskattningar av företagens administrativa kostnader av olika lagar, bl.a. skattelagarna.²⁵

I USA används ett diagnostiskt formulär för att på ett objektivt sätt bedöma komplexiteten i nya lagförslag²⁶. Ett annat mått som har en koppling till fullgörandet är hur många eller stor andel som lämnar uppgifter via e-deklaration.

Ytterligare en idé skulle kunna vara att mäta hur många skattebetalare som överhuvudtaget har haft kontakt med Skatteverket i samband med redovisning eller betalning av sin skatt. Idealt sett så borde, åtminstone vanliga löntagare, slippa ha kontakt med Skatteverket. Detta borde vara ett tecken på att systemen är väl utvecklade och välfungerande, reglerna är enkla samt informationen är bra och lättillgänglig. Ett mått som visar hur många som har haft kontakt med Skatteverket i samband med deklarationen skulle kunna ge information om systemfunktionen. (se vidare avsnitt 5.2)

Samhällsmedborgarens rättigheter

Samhällsmedborgarens rättigheter handlar om att den enskilde kan lita på att skattereglerna efterlevs i ett totalperspektiv.

Tilliten

Liksom för uppdraget som skattebetalare genomförs redan undersökningar om den enskildes uppfattning om Skatteverket och skattesystemet ur ett medborgarperspektiv. Det handlar här om förtroendet i stort för hur Skatteverket utför sitt uppdrag och om hur skattesystemet fungerar i stort. Denna del täcks upp av enkätfrågorna nedan.

- Totalt sett har jag förtroende för skattesystemet
- Vad tycker du i stort om skattesystemet, dvs. skatternas storlek och reglernas utformning?
- Totalt sett har jag förtroende för Skatteverket
- Vad tycker du i stort om skatteverket, dvs. tjänstemännen och det sätt på vilka de utför sina arbetsuppgifter?
- Det är troligt att Skatteverket skulle upptäcka om jag skattefuskade
- Under senare år har skattekontrollen blivit effektivare så att skattefusk i högre grad upptäcks

²⁵ Resultaten av mätningarna av kostnaderna av den administrativa bördan, och förändringarna av dem, redovisas i en databas benämnd Malin. Åtkomst till databasen fås via Nuteks hemsida, www.nutek.se.

²⁶ AICPA Tax Complexity Index, finns översatt till svenska och beskrivet i PM från analysenheten, Annika Persson 2003-09-03

- Skattemyndigheten är bra på att bekämpa skattefusket
- I Sverige betalar alla eller nästan alla den skatt de ska

Även frågan ”För min egen del är det OK att folk skattefuskar om de har möjlighet till det” säger något om tilliten.

Uppfyllelsen

Uppfyllelsen handlar om att hitta objektiva mått på i vilken grad Skatteverket fullgör sitt yttersta uppdrag. Det ideala måttet skulle vara i vilken omfattning Skatteverket tar in den teoretiskt korrekta skatten. Sådana beräkningar har gjorts med ojämna mellanrum i form av t.ex. uppskattningar av den svarta sektorns storlek och olika skattefelsberäkningar. Nyligen har Skatteverket presenterat en omfattande sammanställning av skattefelet och dess delar i den s.k. skattefelskartan.

Problemet med den nämnda typen av uppskattningar är att de kräver stora resurser att genomföra. De är också mycket osäkra eftersom de försöker uppskatta något som definitionsmässigt är dolt. För skattefelsuppskattningar tillkommer dessutom problemet att beskattningen i sig påverkar de aktiviteter som beskattas. Med andra ord så skulle många av de skattepliktiga aktiviteter som man räknar den uteblivna skatten på aldrig ha utförts om full beskattning hade genomförts.

Trots de praktiska och teoretiska problemen är den typ av uppskattningar som nämns ovan värdefulla informationskällor när det gäller att bedöma i vilken grad Skatteverket uppfyller sitt uppdrag och hur verksamheten kan förbättras. Skattefelsberäkningar och dylikt bör därför ingå som en del i effektmätningen med avseende på de övergripande målen.

Ett tänkbart komplement till skattefelskartan skulle vara en karta med det omvända perspektivet, ett medverkandeperspektiv. En karta som koncentrerar sig på de skatter som kommer in och vilka system och mekanismer som ligger bakom skulle möjligen ge ett annat perspektiv på frågan hur Skatteverket ytterst utför sitt uppdrag. En skiss över en sådan s.k. medverkandekarta lämnas nedan (avsnitt 5.2.1).

Skattebetalarens/samhällsmedborgarens skyldigheter

Medborgarens skyldigheter handlar om den enskildes ansvar för att på ett korrekt sätt bidra till finansieringen av de offentliga nyttigheterna. Denna del av det övergripande målet mäts subjektivt genom skattebetalarnas uppgivna lojalitet och objektivt genom hur skattereglerna efterlevs.

Lojaliteten

Bedömningskriteriet *Lojaliteten* fångar in medborgarens skyldigheter ur ett subjektivt perspektiv. Det omfattar i någon mening också Skatteverkets vision om ett samhälle där alla vill göra rätt. Något som alla Skatteverkets aktiviteter ytterst ska sträva mot. Ur mätsynpunkt kan denna del tänkas kunna bedömas genom enkätfrågor som berör frågor om viljan att göra rätt, om man är beredd att undanhålla inkomster från beskattning etc.

Efterlevnaden

Efterlevnaden är den objektiva sidan av skattebetalarens skyldigheter. Ett tänkbart mått på *Efterlevnaden* skulle vara omfattningen av fel i uppgiftslämnandet. Intressant information i sammanhanget skulle dessutom vara om dessa fel kunde delas upp i medvetna och omedvetna. Definitionsmässigt är de medvetna felen den objektiva motsvarigheten till det subjektiva måttet bristande *Lojalitet*. Följaktligen är skattskyldiga i avsaknad av medvetna fel de lojala skattebetalarna. Detta innebär att om det trots hög *Lojalitet* skulle förekomma många fel så beror det på en hög andel omedvetna fel.

I praktiken är dock dessa mått svåra. För det första är det komplicerat att få en uppfattning om felens omfattning överhuvudtaget. För det andra är det i många sammanhang i det närmaste omöjligt att skilja avsiktliga fel från oavsiktliga, eftersom indelningen bestäms av vad som rört sig i huvudet på skattebetalaren. För det tredje är det svårt att ens i teorin definiera vad som är ett avsiktligt respektive oavsiktligt fel. Alla medborgare är skyldiga att känna till alla lagar och regler, så även alla skattebetalare. Det är inget konstigt med det, men hur betraktas i detta sammanhang ett fel som begås av en skattskyldig som på grund av ignorans, lathet eller bristande förståndsgåvor inte brytt sig om att sätta sig in i regelverket? Kan det vara fråga om ett avsiktligt fel även om den skattskyldige säger sig haft viljan men inte förmågan att göra rätt? Kan det vara ett oavsiktligt fel trots att den skattskyldige känt till sin skyldighet att sätta sig in i reglerna för att göra rätt? Innebär detta att det skulle behövas en tredje kategori: ignoransfel?

En indelning av felen i avsiktliga och oavsiktliga är alltså i inte så lätt att åstadkomma. Det bästa som går att göra ur effektmätningssynpunkt är att punktvis genom kontroller försöka få en uppfattning om förekomsten av fel överhuvudtaget och till detta göra mycket grova antaganden om fördelningen mellan avsiktliga och oavsiktliga fel.

5.1.6 Andra politikområden

Det övergripande målet ovan är preciserat utifrån politikområdet skatt, tull och exekution, som i huvudsak utgår från ett statsfinansiellt perspektiv. Utöver denna rent fiskala funktion kan skatterna även ha andra syften. Exempel på detta är när skatterna används som styrmedel för exempelvis miljö-, hälso-, närings- eller fördelningspolitiken. Nedan diskuteras ett möjligt förhållningssätt för det fall att det uppstår ett behov av att i effektmätningssmodellen även beakta sådana vidare syften med beskattningen.

För att beskattningen generellt sett ska ha en styrande effekt gäller:

- 1) att Skatteverket fastställer och tar in den skatt som föreskrivs i lagar och förordningar och
- 2) att skatterna rent tekniskt är utformade på ett sätt som medför att de har den styrande effekt som är avsedd.

När det gäller det första villkoret, att Skatteverket tar in den korrekta skatten, så mäts detta i den måldimension som kallas *Uppfyllelsen* med hjälp av skattefels- och medverkandekartorna.

För att bedöma det andra villkoret krävs helt andra analyser än som normalt är av intresse när Skatteverkets verksamhet diskuteras. I många fall har detta slags analyser också en nära koppling till politisk ideologi och praktik. Med anledning därav har frågor som handlar om de resursstyrande skatternas utformning, effektivitet, funktionssätt etc. lämnats utanför effektmätningssmodellen. För det fall att sådana frågor uppkommer vid ett enstaka tillfälle, t.ex. genom särskilt uppdrag i regleringsbrevet, får metoder och angreppssätt bedömas i det

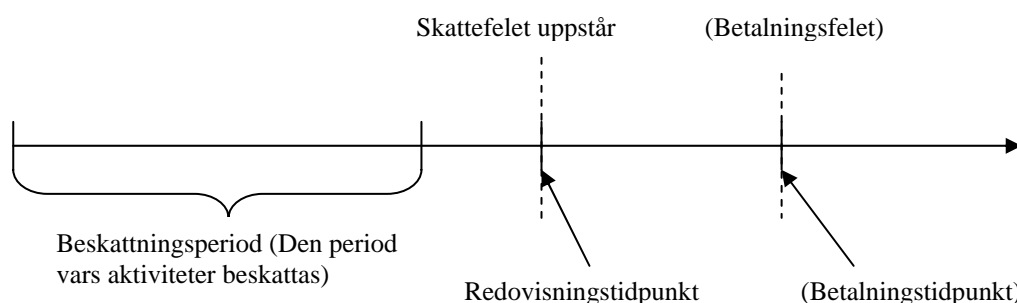
enskilda fallet. Den metodbeskrivning som finns i avsnitt 3 i detta dokument är i sådana fall användbar.

5.2 En tidsindelning av aktiviteterna kan vara utgångspunkt för effektmätning

Ovan har det övergripande målet strukturerats upp på ett ur effektmätningssynpunkt lämpligt sätt. I detta avsnitt kommer intresset att var inriktat mot de olika typer av åtgärder som står Skatteverket till buds i arbetet med att uppfylla målen. De olika typerna av åtgärder kommer att betraktas utifrån hur skattefelet avseende en specifik beskattningsperiod kan påverkas i ett tidsperspektiv, mer om detta nedan.

Skattefelet är inte iakttagbart i sig och därför svårt att uppskatta. Ändå är det något centralt i Skatteverkets verksamhet och något alla i verksamheten måste förhålla sig till. Enligt de uppskattningar av skattefelets storlek som gjorts uppgår det till cirka 133 mdkr²⁷. Frågan här är vad vi menar när skattefelet uppgår till ett visst belopp? Svaret på frågan är att 133 mdkr avser de skatter som borde ha betalats ytterligare om alla under ett år utförda ekonomiska aktiviteter beskattats enligt gällande regler. Detta betyder att det finns ett tidsperspektiv när man talar om skattefelet storlek. Det betyder också att skattefelet uppstår vid en exakt tidpunkt (på motsvarande sätt uppstår betalningsfelet vid en viss tidpunkt).

Fig 5.5 Skattefelets uppkomst för en enskild beskattningsperiod

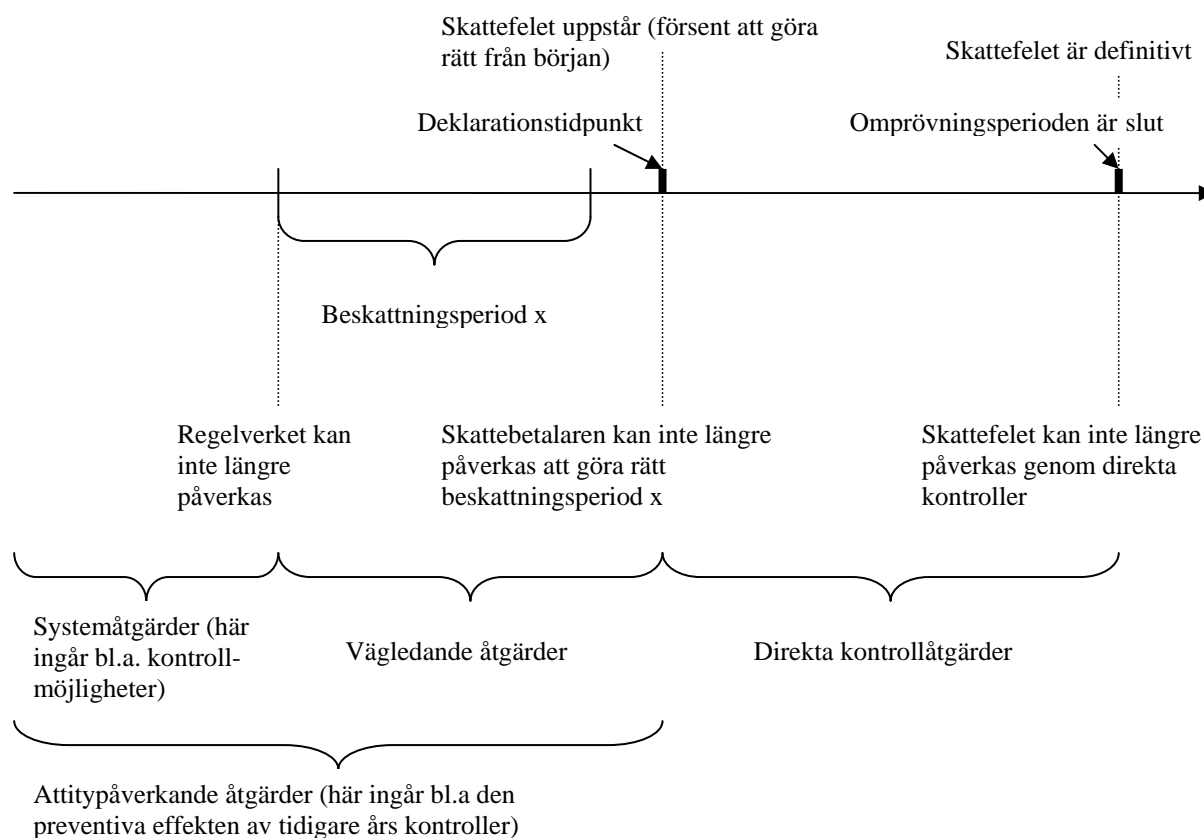


Före den tidpunkt när skatterna ska redovisas finns bara mer eller mindre klara intentioner och avsikter som inte är iakttagbara för någon utomstående. Det är först vid redovisningstidpunkten som de inre avsikterna manifesteras, vilket innebär att det först är då som det är relevant att tala om ett skattefel. Vidare är det viktigt att ha klart för sig att detta skattefel avser de skatter som skulle ha redovisats för en viss beskattningsperiod. Det går därför i strikt mening inte att tala om ett skattefel, som ökar eller minskar över tiden. Istället handlar det om ett nytt skattefel varje år, som är oberoende av tidigare års skattefel.

Om ett års skattefel och tidsindelningen ovan används som utgångspunkt kan Skatteverkets olika typer av åtgärder struktureras upp och kategoriseras efter en tidslinje.

²⁷ Skattefelskartan, SKV 234, och SKV Rapport 2008:1, Skattefelskarta för Sverige

Fig 5.6 De olika åtgärdstypernas påverkan i tiden



Med systemåtgärder avses underhållet och utvecklingen av den infrastruktur som krävs för att skattesystemet ska fungera. I denna infrastruktur ingår allt som krävs för redovisning, betalning och kontroll. Det handlar här om allt från lagstiftning till tekniska system och blankettutformning.

Sett i ett tidsperspektiv kan dessa systemåtgärder sättas in till och med den tidpunkt när beskattningsperioden börjar. Då går gränsen för när en ny lagstiftning kan träda i kraft eller när ett nytt system för redovisning och betalning ska fungera.

Fram till deklarationstidpunkten kan Skatteverket direkt påverka skattebetalarna genom vägledning och attitydpåverkande åtgärder. Vägledning handlar främst om att leda de lojala skattebetalarna rätt, medan attitydpåverkan handlar om att få de ovilliga att vilja göra rätt.

När deklarationstidpunkten har inträffat upphör möjligheterna till dessa mjuka åtgärder. När redovisning av de skattepliktiga aktiviteterna ska ha skett återstår endast kontroll för att direkt påverka skattefelet. Kontroller kan sedan företas fram till den tidpunkt när skattebeslut inte längre kan fattas genom omprövning. Skattefelet för den aktuella perioden är därmed definitivt.

Sett på detta sätt kan skatteverkets åtgärdsparck delas in i fyra delar, *systemåtgärder*, *vägledning*, *attitydpåverkan* och *direkt kontroll*.

Kontrollens preventiva effekt

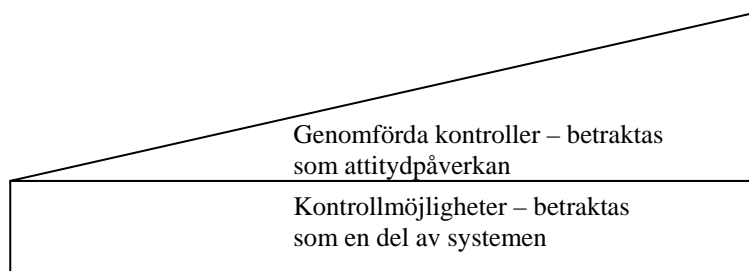
Som konstateras ovan har kontrollen endast en direkt effekt på skattefelet det aktuella året. Sett framåt i tiden har dock kontrollen även en indirekt effekt. Denna indirekta effekt kallas

ofta preventiv effekt, och är detsamma som den drivkraft att göra rätt som den blotta möjligheten av en kontroll skapar hos skattebetalarna.

Detta innebär att de kontroller som genomförs den aktuella beskattningsperioden har en indirekt effekt på skattebetalarna kommande perioder. Utöver den effekt som kan anses härröra från skatteverkets faktiska kontrollaktiviteter kan det även sägas finnas en grundläggande påverkan enbart på grund av de kontrollmöjligheter som finns. Med andra ord så bör det finnas en i princip konstant påverkan, som beror på vilka legala och praktiska kontrollmöjligheter som finns. Denna påverkan kan sedan bli större beroende på vilka kontrollaktiviteter som Skatteverket faktiskt företar.

I tidslinjen ovan betraktas kontrollmöjligheterna som en del av systemåtgärderna, eftersom de främst har med lagstiftning att göra. Den preventiva effekten därutöver, som alltså är beroende av inriktningen på tidigare års kontroller, betraktas som attitydpåverkande åtgärder. Det hela kan illustreras som i figur 5.7 nedan.

Fig 5.7 Kontrollens preventiva effekt delas upp i två delar i effektmättningsmodellen



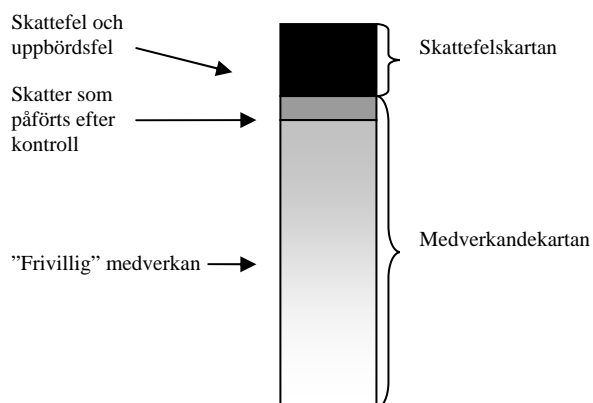
5.2.1 Medverkandekartan

Skatteverket brukar ofta presentera sin bild av de totala skatterna genom ett stapeldiagram. Hela stapeln symboliserar den teoretiskt riktiga skatten. Den översta delen av stapeln symboliserar den skatt som inte redovisas men skulle ha redovisats om alla regler följdes korrekt. Den undre delen visar de skatter som kommer in på "frivillig" väg. Ibland markeras en viss del av den undre delen som skatt som betalas som en följd av preventiv effekt av kontrollen. Mitt emellan de två delarna finns en liten del som symboliserar den skatt som direkt fastställs genom kontrollinsatser.

Skatteverkets analysenhet har under 2007 tagit fram en skattefelskarta²⁸ i vilken uppskattas storleken på och fördelningen av det s.k. skattefelet, dvs. den översta delen av stapeln. I detta dokument diskuteras även möjligheterna att strukturera upp den undre delen av stapeln efter vilka mekanismer som gör att skatterna redovisas, fastställs och betalas, uttryckt i termer av system. Skattefelskartan skulle i så fall kunna kompletteras med en medverkandekarta.

²⁸ Skattefelskartan, SKV 234, och SKV Rapport 2008:1, Skattefelskarta för Sverige

Fig 5.8 Skattefels- och medverkandekartans delar av skattestapeln



En stor del av skatterna redovisas och betalas utan någon större medverkan från skattesubjektets sida. Så är fallet för större delen av skatterna på tjänsteinkomster och avkastning på kapital. Dessa skatter är exempel på väl utvecklade system för redovisning, betalning och kontroll. Visserligen förutsätter dessa system en hög grad av medverkan från tredje man, men de är ändå förhållandevis säkra. Detta beror dels på att det finns inbyggda mekanismer som gör att den som är uppgiftsskyldig inte berikar sig själv på skattesubjektets bekostnad, dels på att kontroller mot uppgiftslämnandet kan utföras effektivt. Ett system som hade byggt på att varje skattskyldig själv redovisade och betalade sin skatt hade varit oerhört mycket osäkrare och ineffektivare.

För andra skatter finns säkra system för fastställande av skatt. Betalningen sköts dock av den skattskyldige själv. Detta gäller främst fastighetsskatten. För andra skatter finns sådana tredjemansuppgifter att redovisning av den skattepliktiga aktiviteten inte går att undvika. Uppgifterna är dock sådana att den korrekta skatten inte går att fastställa utan medverkan från den skattskyldige. Så är fallet för reavinsterna på t.ex. aktier och fastigheter. Därutöver finns för vissa skatter andra indikatorer och mekanismer som stödjer kontrollen och systemen. Exempel på detta är momsens beskattningskedja som i sig innehåller självkontrollerande mekanismer samtidigt som den även ger vissa kontrollindikatorer. Ett annat exempel på kontrollmekanismer är revisionskravet för aktiebolag.

Exemplen ovan visar hur skatter i olika grad kan säkerställas beroende på hur systemen för fastställande, betalning och även externa krav är utformade. Andra skatter är i högre grad beroende av någon slags frivillighet. En ungefärlig bild av hur stor del av de totala skatterna som kommer in tack vare väl utvecklade system respektive på andra sätt skulle kunna ges av en medverkandekarta enligt skissen nedan.

Fig 5.9 Medverkandekartan

			Inkomstskatt, tjänst	Skatt på kapital	Bolags-skatt	Fastighets-skatt	Inkomst-skatt nrv	Sociala avgifter	Moms	Punkt-skatt	Övriga skatter
Direkta kontroll-åtgärder	Skatten fastställd genom kontroll		Viss del	Viss del	Viss del	Viss del	Viss del	Viss del	Viss del	Viss del	Viss del
Vägledande och attitydpåverkande åtgärder	Systemet varken fastställer eller indikerar beskattning	Skatten redovisas med hög grad av "frivillighet"	Avdrag (negativ post)	Liten del		Mycket liten del (utländska bostäder)	Mycket stor del	Viss del (egenavgifter)	Viss del (ink dekl fys)	Viss del	
		Andra kontrollmekanismer finns			Stor del					Viss del	
Systemåtgärder	Ej fastställt genom system men de indikerar att beskattning ska ske	Vissa kontrollindikatorer finns		Viss del					Stor del	Viss del	
		Systemet ger goda kontrollindikationer		Viss del (försäljningar aktier, fastigheter)				Stor del (AG-avg)			
	Fastställt genom systemet		Mycket stor del (löneinkomster)	Stor del (räntor, utdelningar, försäljningar aktiefonder)		Nästan allt				Viss del	

Medverkandekartan är alltså tänkt som en kategorisering av skatteuppbörden efter vilka mekanismer, uttryckta i termer av system, som ligger bakom skatternas fastställande. Det är viktigt att poängtera att medverkandekartan inte ger sig ut för att fullständigt förklara orsakerna till att skatterna fastställs. Den ska istället utgöra en grund utifrån vilken bedömningar av olika åtgärders effekter och diverse analyser kan göras.

Nedan diskuteras närmare vilken betydelse de fyra åtgärdskategorierna har i medverkandekartan.

Systemåtgärder

De skatter som fastställs genom systemen, eller där kontrollmöjligheterna är sådana att de inte går att undvika, är ett resultat av tidigare genomförda systemåtgärder. För dessa skatter är systemåtgärderna fortsatt viktiga, främst som underhåll av befintliga system för att slå vakt om nuvarande uppbörd. Systemåtgärder är alltså viktigare ju närmare basen i skattefelskartan som skatten finns. Systemen har dock även betydelse för skatter som befinner sig längre upp. För dessa skatter handlar det till stor del om kontrollmöjligheter. En förbättring av kontrollmöjligheterna förflyttar skatten nedåt i medverkandekartan. Ett exempel på en sådan nedflyttning är införandet av personliggerare i kontantbranscher.

Vägledande och attitydpåverkande åtgärder

Ju mindre betydelse systemen har för skatternas fastställande, desto viktigare blir de kompletterande åtgärderna vägledning och attitydpåverkan.

För skatter som inte helt fastställs genom systemen fordras en mer eller mindre stor medverkan från skattebetalarnas sida. Ju mer komplicerade reglerna är desto större är kraven på skattebetalarnas insats. I de fall skattebetalarna har en vilja att göra rätt handlar Skatte-

verkets åtgärder främst om vägledning, dvs. om en hjälp att göra rätt. Om Skattebetalarna däremot saknar en vilja att göra rätt kan Skatteverket gå in med attitydpåverkande åtgärder.

Direkta kontrollåtgärder

Mellan alla skatter som fastställs på normalt sätt och de skatter som aldrig fastställs, och därmed ingår i skattefelskartan, finns de skatter som direkt fastställs genom kontroll. Dessa ingår i skattefelskartan och har fått en mörkgrå markering ovan. Eftersom de alla fastställs direkt genom en åtgärd sammanfaller denna med fältet i medverkandekartan.

5.3 Vad som kan mätas av enskilda åtgärder – sammankoppling av åtgärder och måldimensioner

I avsnitt 5.1 delades det långsiktiga målet upp i ett rättighets- och ett skyldighetsperspektiv. Rättighetssidan består av fyra måldimensioner. Dessa är: *Mötet*, *Fullgörandet*, *Tilliten* och *Uppfyllelsen*. Skyldighetssidan har två måldimensioner, *Lojaliteten* och *Efterlevnaden*. I avsnittet ovan indelades Skatteverkets aktiviteter in efter ett tidsperspektiv. Denna indelning gav aktiviteterna: *systemåtgärder*, *vägledande åtgärder*, *attitydpåverkande åtgärder* och *kontrollåtgärder*. I matrisen nedan kopplas varje aktivitet samman med varje måldimension för en bedömning av vilket samband de har, och i förlängningen vad som kan mätas och vad som kan utläsas av måtten.

Fig 5.10 Sammankoppling av åtgärdstyper och måldimensioner

Aktivitet	Måldimensioner					
	Rättigheter				Skyldigheter	
	Skattebetalaren		Samhällsmedborgaren		Skattebetalaren/samhällsmedborgaren	
	Mötet	Fullgörandet	Tilliten	Uppfyllelsen	Lojaliteten	Efterlevnaden
System-åtgärder	Visst samband	Samband	Samband	Samband	Samband	Samband
Vägledande åtgärder	Samband	Samband	Visst samband	Samband	Samband	Samband
Attitydpåverkande åtgärder	Komplicerat samband	Komplicerat samband	Komplicerat samband	Komplicerat samband	Komplicerat samband	Komplicerat samband
Direkta kontroll-åtgärder	Samband	Samband	Inget samband per. def. (ev. följd effekter betraktas som attitydpåverkan kommande perioder)	Direkt samband aktuell period (ev. följd effekter avser kommande perioder)	Inget samband per. def. (ev. följd effekter betraktas som attitydpåverkan kommande perioder)	Inget samband per. def. (ev. följd effekter betraktas som attitydpåverkan kommande perioder)

I matrisen finns alltså 24 kombinationer av åtgärder och måldimensioner. För varje kombination finns i matrisen angivet vilken grad av samband som finns mellan åtgärd och måldimension. Respektive kombination kan sedan användas för två olika slags analyser.

Den första analysen är för uppföljning av hur satsningar på olika typer av åtgärder avspeglar sig i de olika måldimensionerna. Utgångspunkten är alltså de olika åtgärdstypernas effekter på det övergripande målet.

Den andra analysen utgår från hur måluppfyllelsen ser ut utifrån de olika måldimensionernas perspektiv. Denna analys kan sedan användas för att planering av hur de fyra olika åtgärdstyperna kan användas för att förbättra verksamheten och dess resultat.

Nedan beskrivs närmare de två olika typerna av analyser var för sig. Först beskrivs hur modellen kan användas för uppföljning av olika typer av åtgärder, sedan hur den kan användas för strategisk planering av verksamheten.

5.3.1 Den första analysen - uppföljning av åtgärder och mätning av effekter

Nedan förs för varje typ av åtgärd ett kortare resonemang om vilka samband som kan finnas mellan de olika måldimensionerna, och vad som kan utläsas av de olika måtten som finns för varje sådan.

Systemåtgärder

När det gäller systemåtgärder så bör det finnas ett ganska tydligt samband med de flesta måldimensioner. Ändringar i systemen vad gäller kontrolluppgiftsskyldigheten har exempelvis ofta en effekt på den kategori som kallas *Fullgörandet*. Denna effekt kan dock vara positiv för en kategori skattskyldiga medan den kan vara negativ för en annan. Skattskyldiga som genom en utökad kontrolluppgiftsskyldighet får minskade krav på uppgiftslämnande får också minskade fullgörandekostnader, medan effekten blir den motsatta för den part som åläggs uppgiftsskyldigheten. När det gäller *Tilliten* bör ändringar i systemen också ha en effekt. Förbättringar i systemen vad gäller fastställande, betalning eller kontroll bör ha en positiv effekt på samhällsmedborgarnas syn på Skatteverkets möjligheter att fullgöra sin uppgift. På motsvarande sätt har systemändringar en effekt på måldimensionen *Uppfyllelsen*. Åtgärder på detta område visar sig i både medverkande- och skattefelskartan.

När det gäller måldimensionen *Mötet* är eventuella samband med de aktuella måtten mera komplicerat. *Mötet* handlar om hur skattebetalaren subjektivt upplever sina möjligheter att uppfylla sina skyldigheter. Området mäts genom enkätfrågor om t.ex. hur skattebetalaren blivit bemött av Skatteverket, om det är lätt att göra rätt etc. Det som gör sambandet mellan systemåtgärder och *Mötet* komplicerat är att framgångsrika åtgärder på området gör vissa av måtten irrelevanta. Om t.ex. systemen skulle byggas ut i en omfattning som gör att många skattskyldiga kan fullgöra sina skyldigheter utan kontakt med Skatteverket överhuvudtaget så innebär det samtidigt att enkätfrågor om bemötande inte längre är relevanta. Å andra sidan bör förbättringar på detta område avspeglade sig i uppfattningen om det är lätt att göra rätt.

Vägledande åtgärder

När det gäller vägledande åtgärder så bör det finnas ett samband med exempelvis *Mötet* och *Uppfyllelsen*. Hur Skatteverket sköter denna del bör visa sig i enkätfrågor rörande bemötande etc. När det gäller uppfyllelsen så bör god vägledning visa sig genom en minskning av de oavsiktliga felen, vilket på olika sätt ger resultat i skattefels- och medverkandekartorna.

När det gäller vägledande åtgärder och *Tilliten* finns mycket som talar för att det endast är ett svagt samband som råder. De vägledande åtgärderna är främst inriktade mot att avhjälpa omedvetna fel som begås av skattebetalare som har en vilja att göra rätt. Detta är naturligtvis inte oviktigt för den totala *Tilliten* till Skatteverkets förmåga att ta in den korrekta skatten. Det finns dock även många andra faktorer som i sammanhanget är minst lika viktiga. Exempel på sådana kan vara kraftfulla åtgärder mot skattskyldiga som medvetet fuskar eller kanske att verka för enklare regler, vilket i förlängningen skulle minska behovet av vägledning.

Attitydpåverkande åtgärder

Attitydpåverkande åtgärder innefattar många olika typer av åtgärder. Den viktigaste åtgärden med en attitydpåverkande effekt är tidigare års kontroller. Det handlar alltså om den effekt som uppstår som en följd av de kontroller som tidigare genomförts, dvs. kontrollens preventiva effekt. Andra tänkbare åtgärder som har en attitydpåverkande effekt är informationskampanjer.

När det gäller sambandet mellan attitydpåverkande åtgärder och de olika måldimensionerna kan det generellt sägas vara komplicerat, i den meningen att följderna av en insats är svåra att förutsäga. Effekterna är i regel beroende av många andra faktorer som kan variera beroende på omständigheter.

En svårighet med preventiva effekter är att det mellan kontrollinsatsen och målet, uttryckt t.ex. som minskat skattefel, finns en hel kedja med mellanliggande effekter som alla var och en är beroende av många faktorer. Det går därmed inte att säga att en kontroll alltid leder till en viss effekt. Det finns en oerhörd mängd tänkbara utfall för samma insats beroende på omständigheter (mer om detta i avsnitt 6.1.2). Problemet ur effektmätningssynpunkt är att isolera effekten av insatsen från effekten av alla andra faktorer.

En annan svårighet med preventiva effekter är att de måste definieras och avgränsas, både i tid och rum. Tidsaspekten handlar om effektens verkan framåt i tiden. Avses med preventiv effekt ett år framåt i tiden, tio år eller den samlade effekten för evig tid framåt? Rumsaspekten handlar om hur den avgränsas i mängden den ska verka på. Den individualpreventiva effekten handlar om dem som kontrolleras, men hur avgränsas den allmänpreventiva effekten? Till dem inom samma kategori som kontrollerats men inte själva utsatts för kontroller? Till hela skattebetalarkollektivet? Avser man effekten inom den företeelse som kontrollerats eller påverkan på hela skattefelet? Om man avser hela skattefelet: hur tar man hänsyn till eventuella negativa effekter i form av ökat fel inom andra områden som inte kontrolleras?

I nästa avsnitt förs ett kortare resonemang om hur dessa försvårande faktorer i mätningen av kontrollens preventiva effekter kan hanteras med ett flerårigt cykliskt perspektiv.

Även informationskampanjer med attitydpåverkande syfte är svåra ur effektmätningssynpunkt. Professionellt utförda kampanjer bör ha en effekt. Exempel på mer problematiska områden kan dock vara när Skatteverket i syfte att informera om framgångar inom kontrollverksamheten ger en bild av att stora skattebelopp undanhålls inom vissa områden eller branscher. Hur påverkar en sådan negativ bild allmänhetens tillit till Skatteverket och skattesystemet eller deras lojalitet till en skattefinansierad offentlig sektor?

Direkta kontrollåtgärder

Direkta kontrollåtgärder har i den aktuella beskattningsperioden ett direkt samband med måldimensionen *Uppfyllelsen*. Med direkt samband menas att en åtgärds påverkan direkt kan avläsas i något mått. Det innebär också att det är åtgärden och inget annat än just denna åtgärd som åstadkommit den iakttagna förändringen. *Uppfyllelsen* är det objektiva måttet på hur Skatteverket utför sitt uppdrag. Denna måldimension mäts genom skattefelskartan och medverkandekartan. I och med att kontrollresultaten ingår i medverkandekartan finns också ett direkt samband mellan direkta kontrollåtgärder och uppfyllelsen.

Direkta kontrollåtgärder har också ett samband med måldimensionerna *Mötet* och *Fullgörandet*. En kontrollerad skattebetalare bör efter en kontroll ofta ha någon uppfattning om bemötandet, vilket kan visa sig i en enkätfråga. Likaså är kontroller ofta förenad med någon kostnad för skattebetalaren, dvs. kontroller påverkar i regel fullgörandekostnaderna.

När det gäller övriga måldimensioner så saknar de direkta kontrollåtgärderna samband med dessa för den aktuella beskattningsperioden. Eftersom felet redan är begånget kan åtgärden inte ge någon annan effekt än just den direkta effekten. Eventuella följd effekter av kontrollen på *Tilliten*, *Lojaliteten* eller *Efterlevnaden* betraktas, liksom den preventiva effekten, som attitydpåverkan och avser kommande beskattningsperioder. Denna aspekt berörs nedan.

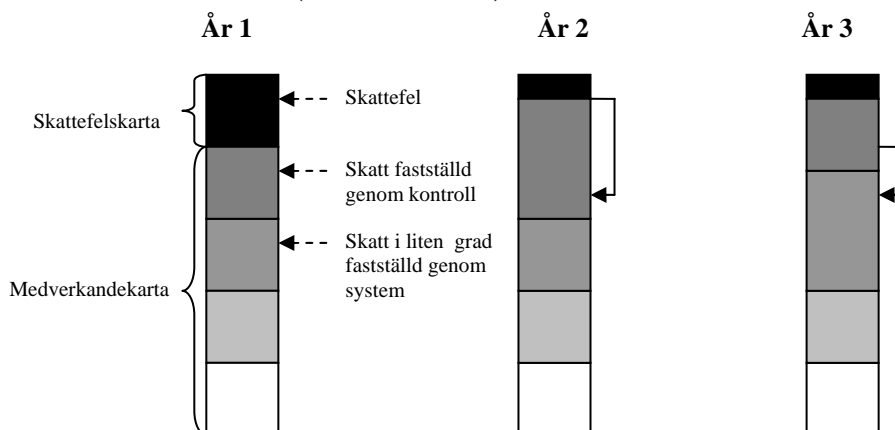
5.3.2 Utveckling av den första analysen – följd effekter visar sig i ett längre perspektiv

Preventiv effekt

Som beskrivits ovan är kontrollens preventiva effekt svår att hantera ur effektmätningssynpunkt. Detta eftersom en viss periods kontrollåtgärder har en direkt effekt samma period, medan eventuella indirekta effekter kommer först i senare perioder. Med andra ord har kontrollen följd effekter som ur mätsynpunkt måste hanteras i ett längre tidsperspektiv än en beskattningsperiod. Den beskrivna enperiodsmodellen räcker här inte till för att beskriva kopplingen mellan insats och effekt. Den måste därför betraktas i ett flerperiodiskt, eller cykliskt perspektiv, för att alla effekter ska fångas in.

Rent teoretiskt skulle kontrollens indirekta effekt avspeglas i kontinuerliga uppdateringar och analyser av skattefels- och medverkandekartorna. Den direkta effekten av kontrollen skulle visa sig i en minskning av skattefelet och en samtidig ökning av de direkta kontroll-effekterna (år 2). De dynamiska effekterna skulle bestå i en minskning av de direkta kontrollåtgärderna och en samtidig ökning av "frivillig" medverkan (år 3) (allt annat lika).

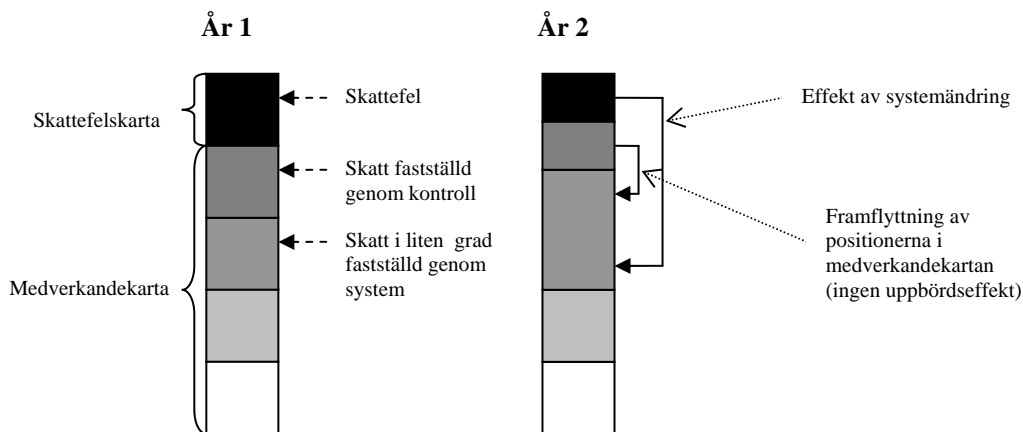
Fig 5.11 Kontrollens dynamiska effekter visad genom tre års jämförelser av skattefels-/medverkandekartorna (allt annat lika)



Den ovan beskrivna preventiva effekten utgår från att de kontroller som faktiskt genomförs har en effekt framåt i tiden. Ofta talas dock om att kontrollens blotta möjlighet även den har preventiv effekt, dvs. många redovisar sin skatt på ett korrekt sätt enbart för att Skatteverket kan genomföra en kontroll. Det handlar alltså om den uppdelning av preventiv effekt som beskrevs i avsnitt 5.2. I effektmätningssmodellen betraktas kontrollmöjligheterna, dvs. de ramar regelverket ger, som en del av systemen, medan de faktiskt genomförda kontrollerna hör till de attitydpåverkande åtgärderna.

En förändring i kontrollmöjligheterna, tex. införande av personalliggare, skulle i medverkande-/skattefelskartorna visa sig som nedan.

Fig 5.12 Förändrade kontrollmöjligheter visad genom två års jämförelser av skattefels-/medverkandekartorna (allt annat lika)



Resonemanget ovan om hur kontrollens preventiva effekter manifesteras i skattefels- och medverkandekartorna är teoretiskt. I praktiken finns stora mätproblem, inte minst vad gäller isolerbarhet. De praktiska problemen vad gäller detta behandlas vidare i kapitel 6 och i avsnitt 3. I avsnitt 5.3.4 diskuteras dock kontrollens preventiva effekter i medverkande- och skattefelskartan ytterligare något.

Samband mellan måldimensioner

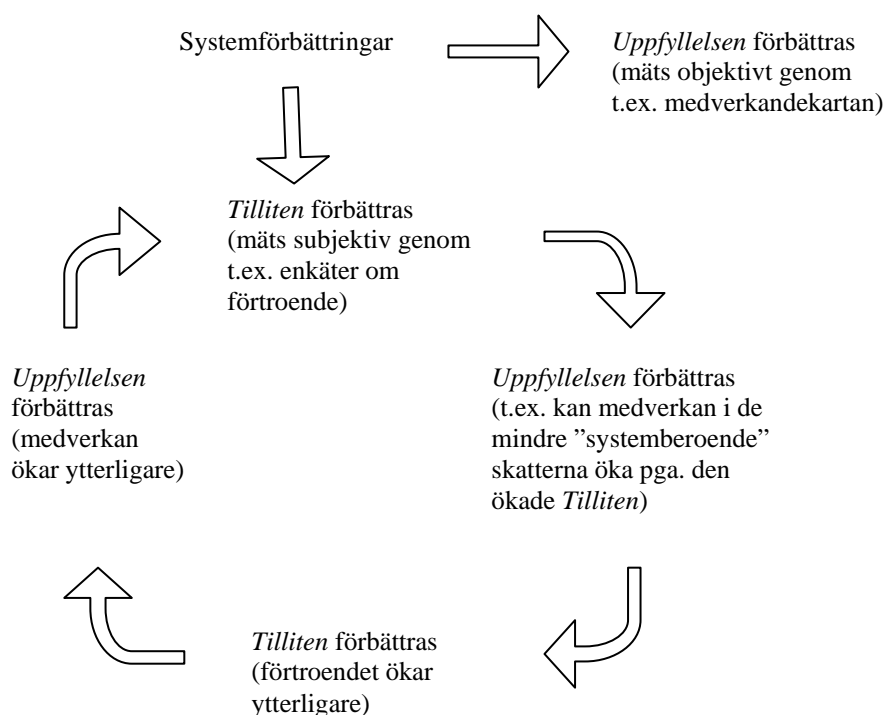
Resonemanget ovan har varit begränsat till vilka eventuella effekter olika typer av insatser har på de olika måldimensionerna, betraktade en i taget. Måldimensionerna betraktas som olika aspekter av ett och samma övergripande mål. Det kan dock vara mer komplicerat än så. Förutom att vara mål i sig kan det även tänkas att måldimensionerna är medel för uppnåendet av andra mål. En insats kan exempelvis få en effekt som är mätbar i två måldimensioner, en objektiv och en subjektiv. Förändringar i den objektiva dimensionen kan dock i sig tänkas få en effekt på den subjektiva, vilket kan leda till en självförstärkande effekt.

Ett exempel kan göra resonemanget tydligare. I effektmätningens modellen betraktas den del av Skatteverkets uppgift som går ut på att i så hög omfattning som möjligt ta in den demokratiskt beslutade skatten som ett uppdrag från alla samhällsmedborgare. I vilken grad Skatteverket utför detta uppdrag mäts subjektivt i den måldimension som kallas *Tilliten* och objektivt i den som kallas *Uppfyllelsen*. En hypotes i detta dokument är att effektiva och välfungerande system hjälper Skatteverket i detta sitt uppdrag. Förbättringar på området skulle därmed också visa sig i båda de nämnda måldimensionerna. Att låta analysen stanna där skulle innebära att man enbart såg på systemens funktion isolerat från allt annat. Systemen är dock beroende av många andra faktorer för att fungera, t.ex. ett grundläggande förtroende från alla medborgare och företag för Skatteverket som institution. Givet ett visst förtroende kan alltså förbättringar i systemen förbättra skattesystemets funktionssätt, vilket visar sig i mätningar av de subjektiva och objektiva målen.

Eftersom förtroendet för Skatteverkets som institution både är en viktig faktor för systemens funktionssätt och ett mål i sig i måldimensionen *Uppfyllelsen*, blir det både ett *mål* och ett *medel*. Förtroendet är alltså ett mål i sig samtidigt som det är ett medel för att nå andra mål.

Förändringar i systemen får med andra ord följd effekter som är svåra att mäta och överblicka. En tänkbar kedja visas i figur 5.10 nedan.

Fig 5.13 Måldimensionerna kan vara både mål och medel



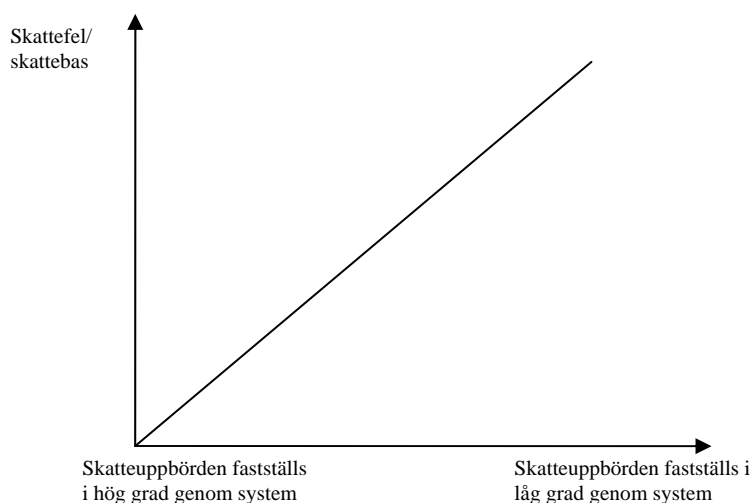
Förutom att vara direkta mått på effekterna av en systemåtgärd kan både *Uppfyllelsen* och *Tilliten* vara medel i en självförstärkande process.

5.3.3 Den andra analysen – hur kan modellen användas för att hitta förbättringsmöjligheter i verksamheten?

Den andra typen av analys utgår från resultaten i en viss måldimension. Tanken är att analysen ska kunna användas för att dra upp åtgärdsstrategier eller bedömningar av inom vilka områden olika insatser har bäst effekt. Nedan ges ett exempel på hur verksamheten kan analyseras utifrån den information som finns i skattefels- och medverkandekartorna.

När medverkandekartan betraktas tillsammans med skattefelskartan går det att få en uppfattning om hur stort skattefelet är i förhållande till de skatter som fastställs, dvs det går att få en uppfattning om det relativa skattefelets storlek för de olika skatteslagen. En rimlig hypotes är att det finns ett negativt samband mellan grad av systemutbyggnad och det relativa skattefelet. Exempelvis bör det relativa skattefelet för tjänsteinkomster, som i hög grad beskattas genom väl utbyggda system för fastställande och kontroll, vara betydligt lägre än för inkomster från näringsverksamhet, som i mycket lägre grad fastställs genom system. Hypotesen illustreras i diagrammet nedan (Fig 5.12, diagrammet har lånats från SKV Rapport 2008:1, Skattefelskartan i Sverige).

Fig 5.14 Hypotes om samband mellan relativt skattefel och skatteuppbördens systemens roll i skatteuppbörden



Källa: SKV Rapport 2008:1 Skattefelskartan i Sverige

Om hypotesen stämmer bygger det svenska skattesystemet i mycket hög grad på väl utbyggda system. Systemen skulle alltså utgöra en viktig del av fundamentet för den offentliga sektorns finansiering.

Utöver systemen finns dock många faktorer som också de har betydelse för skattesystemets funktionssätt. En faktor som ofta framhålls är förtroendet. Förtroendet för Skatteverket är mycket viktigt, som nämndes ovan inte bara som mål utan även som medel. Fanns inte ett grundläggande förtroende för Skatteverket som institution skulle drivkrafterna att delta och ingå i systemet vara dåliga, vilket väsentligt skulle försvåra möjligheterna att ta in den demokratiskt beslutade skatten. Av allt att döma ligger förtroendenivån för Skatteverket, liksom för de flesta svenska institutioner och för samhället i stort, på en ganska hög och stabil nivå internationellt sett.²⁹ Det hänger sannolikt i hög grad samman med att Sverige är en stabil demokrati, med låg grad av korruption och väl utvecklade effektiva system. Detta är viktiga faktorer som i ett samhällsperspektiv är fundamentala. I just den här diskussionen konstanthålls dock dessa faktorer för ett resonemang om hur skattesystemet kan förbättras genom förändringar av systemen och övriga åtgärder.

Resonemanget ovan bygger alltså på att väl utbyggda system utgör en viktig grund för beskattningen. Detta på grund av att utrymmet för en felaktig beskattning blir mindre när systemen för redovisning, betalning och kontroll är välutvecklade och välfungerande.

Om systemen nu utgör en viktig grund för beskattning: när kommer då Skatteverkets övriga åtgärder in i bilden?

När systemen inte utgör det goda stödet för en effektiv och säker beskattning blir övriga åtgärder viktigare. När systemen kärvar, t.ex. när reglerna är så krångliga att de lojala skattebetalarna behöver stöd och hjälp för att göra rätt, blir de vägledande åtgärderna viktigare. När skattebetalarna inte är samarbetsvilliga i nödvändig omfattning blir de attitydskapande åtgärderna viktigare, dvs. det handlar här om att genom olika typer av

²⁹ Det kan i detta sammanhang nämnas att de nordiska länderna i forskning om social tillit framhålls som närmast extrema exempel på när tilliten till vänner, grannar, politiker, myndigheter etc. är hög, se t.ex. Delhey & Newton, 2004, s. 14 f.

åtgärder göra illojala skattebetalare till lojala sådana. I sista hand när alla andra typer av åtgärder är uttömda återstår den sista typen av åtgärd – direkta kontrollåtgärder.

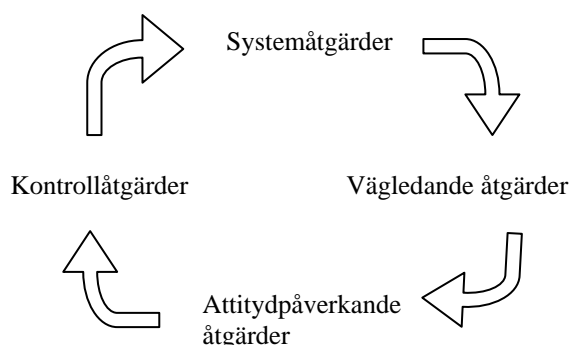
Resonemanget leder fram till åtgärdshierarkin nedan.

- 1) Systemåtgärder
- 2) Vägledande åtgärder
- 3) Attitydpåverkande åtgärder
- 4) Direkta kontrollåtgärder

5.3.4 Utveckling av den andra analysen - ett cykliskt betraktelsesätt av åtgärdstyperna kan ge ytterligare information och kunskap

I avsnittet ovan föreslås alltså att de fyra åtgärdstyperna betraktas hierarkiskt, så att väl utbyggda system utgör en god och förhållandevis stabil grund för beskattning. Först därefter följer övriga åtgärder med de direkta kontrollåtgärderna som en sista uppsamlingsinsats. Samtidigt konstateras i avsnitt 5.3.2 att kontrollens framåtsyftande preventiva effekter är komplicerade ur effektmätningssynpunkt. Kontrollens eventuella attitydpåverkan ett kommande år är med andra ord svår att mäta. Resonemanget nedan, om ett cykliskt och flerårigt betraktelsesätt av åtgärdstyperna, kan dock vara en hjälp att få ytterligare information för den strategiska planeringen av verksamheten.

Fig 5.15 Cykliskt betraktelsesätt av åtgärdstyperna



Antag att det finns ett problemområde med ett konstaterat stort skattefel. Om det efter massiva kontrollinsatser efter några år konstateras en minskad felfrekvens kan det vara en effekt av kontrollen. Man kan då säga att kontrollerna i förlängningen lett till en attitydpåverkan som i sin tur har minskat skattefelet på området. Denna följeffekt av kontrollen beskrivs ovan i avsnitt 5.3.2. Om det däremot efter några år konstateras att skattefelet fortfarande är stort kan det finnas anledning att se efter andra alternativ.

Enligt resonemanget ovan är det effektivaste sättet att säkerställa beskattningen genom förbättringar av systemen. En möjlig systemförändring är utökning av kontrollmöjligheterna. Exempel på sådana förändringar är vidgning av kontrolluppgiftsskyldigheten eller införande av andra kontrollinstrument, t ex personalliggare. Andra typer av systemåtgärder kan vara att på olika sätt underlätta för skattebetalarna att göra rätt. Ett exempel på det senare är Skatteverkets förslag att skapa nya internetbaserade system för uppgiftslämnande och lönebetalningar för arbetsgivare, skatta.se.³⁰

³⁰ SKV Rapport 2007:1, Svartköp och svartjobb i Sverige

Systemförändringar kan dock även ses i ett större perspektiv. I vissa fall kan ett stort skattefel bero på brister i regelverket i sig. Lagregler som inte överensstämmer med allmänt accepterade skatterättsliga principer eller det allmänna rättsmedvetandet kräver i regel stora resurser hos den som har till uppgift att tillse efterlevnaden av dem. I längden är de antagligen svåra att upprätthålla överhuvudtaget. Skatteverket styr inte över regelutformningen. Skatteverket har dock en viktig uppgift att slå larm när det finns tecken på att reglerna inte fungerar och att då komma med förslag på förbättringar och ändringar. Ett aktuellt exempel på en proaktiv handling från Skatteverkets sida är förslaget till en ökning av gränsen för kontrolluppgiftsskyldighet när en privatperson anlitar en annan privatperson. De nuvarande reglerna har bedömts inte överensstämma med det allmänna rättsmedvetandet. Ersättningar för vissa mindre tjänster, som enligt nuvarande lagstiftning medför samma skyldigheter som för en arbetsgivare, uppfattas av det stora flertalet som så bagatellartade att de överhuvudtaget inte funderar över eventuella beskattningskonsekvenser. Skatteverkets förslag kan därmed kallas för en systemanpassning, som i längden bör ha en positiv inverkan på skattefelet.

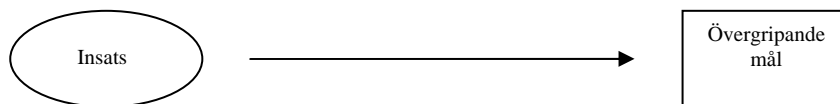
När möjligheterna till systemförändringar är uttömda återstår de mjuka åtgärderna, vägledning och attitydpåverkan, samt i sista hand direkta kontroller. Cirkeln är sluten.

5.4 Sammanfattning och slutsats kapitel 5

I detta kapitel strukturerades det övergripande målet upp i olika mätbara enheter, så kallade måldimensioner. Dessa kopplades sedan samman med Skatteverkets åtgärdsplan, som definieras utifrån hur de i tiden påverkar beskattningen för en enskild beskattningsperiod. Sammankopplingen bedöms kunna ge information för dels olika slags analyser rörande uppfyllelsen av de övergripande målen, dels ge information för olika strategiska analyser. De fyra åtgärdstyperna kan också betraktas hierarkiskt, på så sätt att systemåtgärder är det effektivaste sättet att upprätthålla en stabil beskattning, därefter följer vägledande och attitydpåverkande åtgärder. Den sista möjligheten att påverka skattefelet är genom direkta kontrollåtgärder. Kontrollens preventiva effekter är komplicerade ur effektmätningssynpunkt. Dessa kan i bästa fall iaktas indirekt i ett flerårigt cykliskt perspektiv.

6 Operativt perspektiv - vad kan mätas om utgångspunkten är enskilda insatser?

I detta kapitel intas det omvända perspektivet jämfört med kapitel 5. Utgångspunkten är enskilda aktiviteter och insatser och deras koppling till det övergripande målet.



Nedan behandlas först kopplingen mellan insats och mål, kausaliteten. Därefter följer ett avsnitt om målet och dess mätbarhet i olika aspekter.

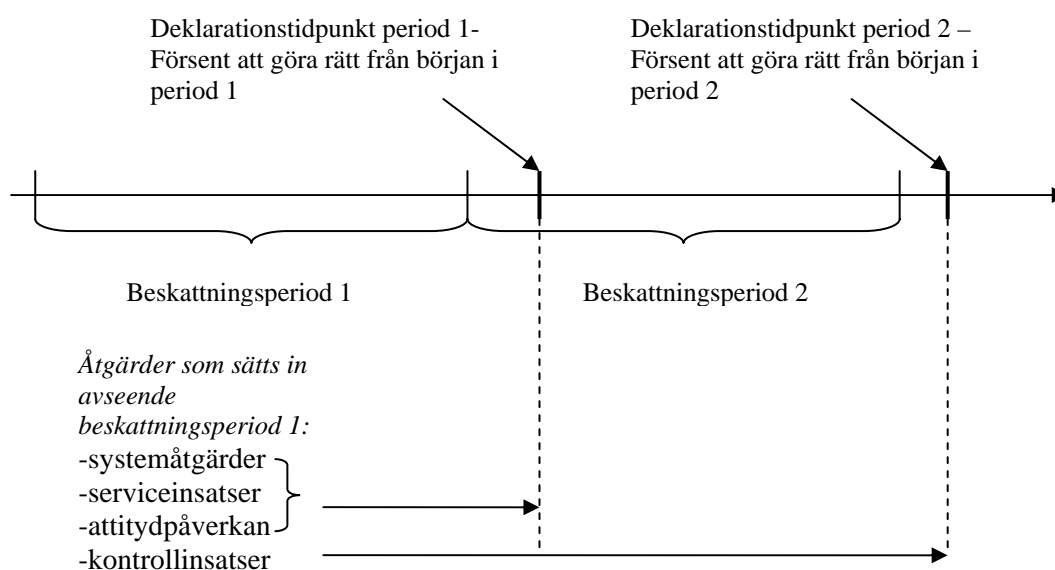
6.1 Kausaliteten

En effekt är en förändring som sker av en orsak. Det ska alltså finnas ett orsakssamband mellan insatsen och den iakttagna förändringen; om inte insatsen hade företagits skulle med andra ord inte förändringen ha skett.

6.1.1 Olika insatserns påverkan i tiden

Om Skatteverkets åtgärdsparck delas in i de fyra kategorierna *systemåtgärder*, *serviceinsatser*, *attitydpåverkan* och *kontrollinsatser*, samtidigt som utgångspunkten är strategin att de skattskyldiga ska förmås att göra rätt från början, så går det att konstatera att insatserna avseende ett visst beskattningsår får effekter på olika beskattningsår framåt i tiden. I och med att det huvudsakligen är kontrollens indirekta effekt, att förmå skattebetalarna att göra rätt från början genom den upplevda risken för upptäckt, som har betydelse i Skatteverkets strategi, har kontrollen främst en effekt i beskattningsperioden efter den att kontrollen satts in.

Fig 6.1 Åtgärdernas påverkan i tiden



Sambanden mellan Skatteverkets olika insatser och deras effekter är således komplexa sett ur ett tidsperspektiv och därmed också ur ett kausalitets- och effektmätningperspektiv. Detta

gäller särskilt kontrollens preventiva effekter eftersom den kommer först i perioden efter den där insatsen sätts in.

6.1.2 Går det att teoretiskt fastställa ett samband mellan en kontrollinsats och det långsiktiga målet?

När det gäller kontrollen så är det en hel kedja som binder samman insatsen med det tänkta slutliga målet, att minska skattefelet. Det tänkta förloppet är att kontrollen ska påverka beteendet inte bara hos den som kontrolleras utan även i en större krets genom s.k. preventiv effekt. En preventiv effekt sker även genom en beteendeförändring i framtiden. En intressant fråga i sammanhanget är om det teoretiskt går att konstatera att det hela alltid förlöper på detta sätt. I sådant fall vore effektmätningen en lätt uppgift. Man maximerar helt enkelt insatsen, ju mer desto bättre.

För att undersöka om så är fallet måste orsakssambanden klarläggas. Insatsens påverkan i tid och rum kan beskrivas genom en matris som delar upp effekten i vilka som påverkas och när detta sker i tiden.

Fig 6.2 Kontrollens effekt i tid och rum

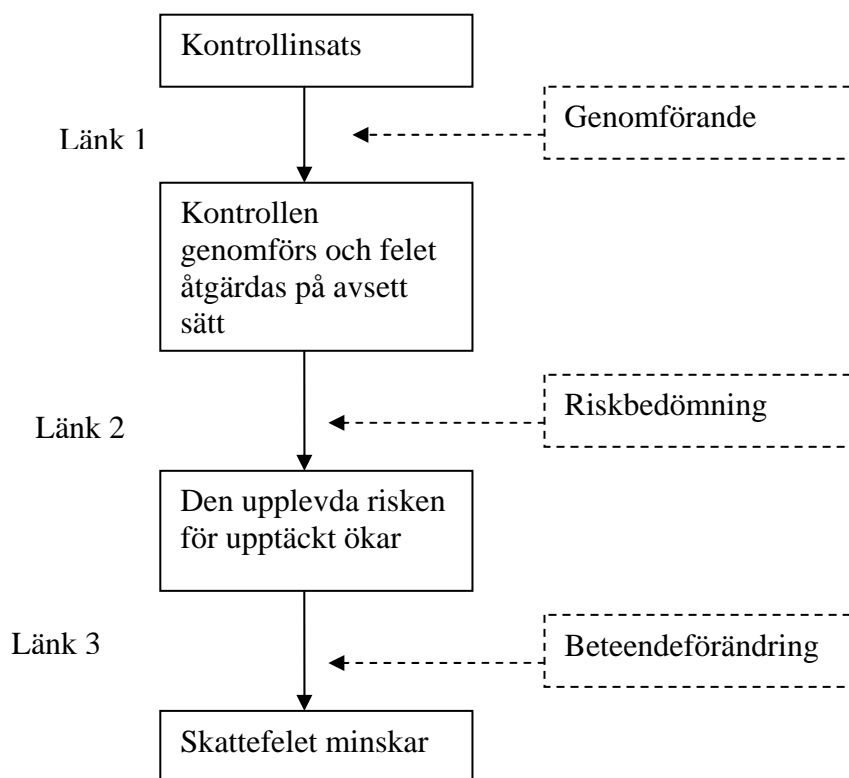
	Den kontrollerade	Alla andra skattebetalare
Aktuellt år	Direkt effekt	Allmänpreventiv effekt*
Kommande år	Individualpreventiv effekt	Allmänpreventiv effekt

*En allmänpreventiv kan tänkas uppstå även under det år kontrollen genomförs om Skatteverket aviserat kontrollinsatsen i förväg

Sambandet mellan kontrollinsatsen och den direkta effekten på den kontrollerade det aktuella året är lätt att konstatera eftersom åtgärden och dess effekt är direkt iakttagbara. Den individualpreventiva effekten är däremot inte direkt iakttagbar. Hur individen beter sig efterföljande år skulle gå att undersöka genom en uppföljningskontroll. En sådan säger dock inget om *vad* som påverkat beteendet. Utöver kontrollen kan det ha funnits andra faktorer som också påverkat beteendet. Denna problematik behandlas utförligare nedan i avsnitt 6.2.4 Problemet med andra påverkansfaktorer blir inte mindre för den som försöker sig på att mäta den allmänpreventiva effekten.

Ett sätt att ändå få en uppfattning om det finns ett kausalt samband mellan insats och effekt är att rita upp en effektkedja. En effektkedja visar alla mellanliggande effekter och de länkar som binder dem samman. Om det empiriskt eller på annat tillförlitligt sätt går att fastställa att varje länk har ett kausalt samband med den föregående kan man också konstatera att det finns ett kausalt samband mellan insats och effekt. Nedan visas två tänkbara effektkedjor, en för den individualpreventiva effekten och en för den allmänpreventiva effekten. Rutorna med de prickade ramarna beskriver länken, dvs den faktor som är avgörande för om det finns ett kausalt samband. Effektkedjorna nedan är enklast möjliga och beskriver ett idealiskt förlopp när ett avsiktligt fel föreligger. Effektkedjor går att göra mycket mer komplicerade med alternativa förlopp och förgreningar.

Fig 6.3 Effektkedja mellan kontrollinsats och minskat skattefel genom individualpreventiv effekt



Länk 1 Genomförande

Det första steget i kedjan är att kontrollen genomförs och att eventuella fel åtgärdas på avsett sätt. Många faktorer är av betydelse i detta sammanhang. Alla förekommande fel hos den skattskyldige bör hittas. Kommunikationen mellan Skatteverket och den skattskyldige måste fungera, bl.a. för att beslutet ska vara rättsligt hållbart, för att den skattskyldige ska få klart för sig vilka regler som gäller etc. Sanktionsavgifter måste påföras på ett korrekt sätt så att det står klart för den skattskyldige att det är kännbart att lämna felaktiga uppgifter.

Länk 2 Riskbedömning

Hur den enskilde skattskyldige bedömer risken för framtida kontroller och att dessa leder till att eventuella felaktigheter upptäcks är beroende av många faktorer. Tre viktiga faktorer är vilka omständigheter som utlöst kontrollen, den tidigare kontrollfrekvensen och om felet är medvetet eller omedvetet. Om omständigheten som utlöst kontrollen är en företeelse som inte funnits tidigare hos den skattskyldige och Skatteverket genast fångar upp detta borde rimligen den skattskyldiges uppfattning om upptäcktsrisken påverkas positivt. Om däremot företeelsen funnits hos den skattskyldige i årtal utan att någon hänt förrän nu, samtidigt som denne hela tiden levt i tron att den gjort rätt, kan uppfattningen om upptäcktsrisken bli den motsatta. Om den skattskyldige blir föremål för kontroll i stort sett varje år förblir uppfattningen om upptäcktsrisken i stort sett opåverkad.

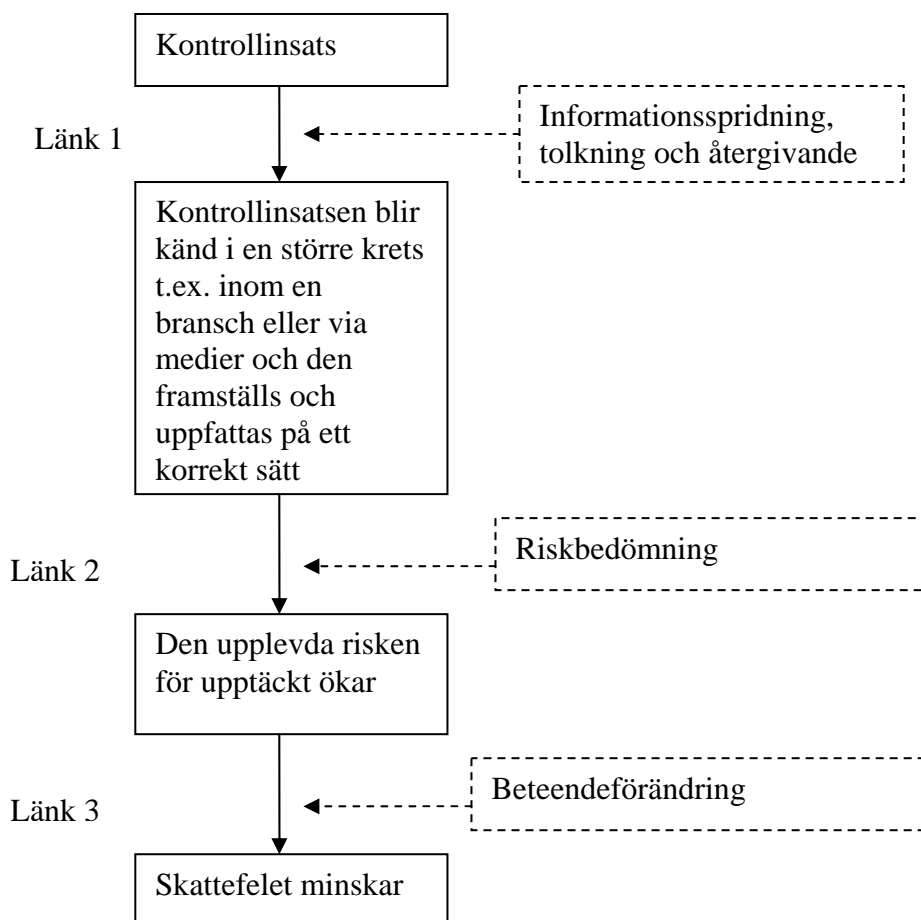
Länk 3 Beteendeförändring

Under förutsättning att den upplevda risken för upptäckt är hög och att denna upptäckt ger kännbara konsekvenser så bör denna länk vara hållbar.

Hela effektkedjan

Det finns en hel del oklarheter och osäkerheter i kedjan. Framförallt hur den skattskyldige reagerar på kontrollen. Av allt att döma finns det många ”om och men” som gör att det inte säkert går att säga om en enskild kontrollinsats leder till att den upplevda risken för upptäckt ökar. Denna osäkerhet gör att det inte går att säga att en kontroll alltid leder till minskat skattefel genom individualpreventiv effekt. Vilka förutsättningar som ska vara uppfyllda för att en kontroll ska ha önskad effekt är ett område som det av allt att döma fortfarande finns stora oklarheter inom.

Fig 6.4 Effektkedja mellan kontrollinsats och minskat skattefel genom allmänpreventiv effekt



Länk 1 Informations-spridning, tolkning och återgivande

Den allmänpreventiva effekten bygger på att insatsen blir känd i en krets utanför den som kontrollerats. Ju vidare krets som får kännedom om insatsen, desto större är den potentiella effekten. Tänkbara medier är bransch- och intresseorganisationer eller massmedier. Ryktesspridning från mun till mun kan också förekomma. Betydelsen av sådan informations-spridning varierar troligen stort beroende på t.ex. hur många potentiellt berörda som finns. Ytterligare en faktor är att insatsen både framställs av avsändaren och uppfattas av mottagaren på ett så korrekt sätt som möjligt.

Denna länk är alltså beroende av många faktorer, t.ex. om den anses ha ett nyhetsvärde bland massmedier, om det finns intresseorganisationer som ser en anledning att informera om saken, om kontrollen blir ett samtalsämne etc. Vidare är det av vikt att den information Skatteverket

lämnar används och återges på ett sådant sätt att bilden av insatsen inte förvanskas eller vinklas gentemot mottagarna.

Länk 2 Riskbedömning

Det är viktigt att poängtera att nästa steg i kedjan handlar om den upplevda risken för upptäckt, vilket är något helt annat än den faktiska risken för upptäckt. Även om Skatteverket inriktar sin verksamhet på ett sätt som gör att den faktiska risken för upptäckt ökar behöver inte den upplevda risken öka. Allt beror på vilken information den skattskyldige får. För att inte belasta resonemanget med detta problem antas nedan att det finns en total överensstämmelse mellan den information Skatteverket släpper och den som når allmänheten.

Även om Skatteverket hade total kontroll över den information som når allmänheten råder stor osäkerhet om vad som påverkar den upplevda risken för upptäckt. Antag t.ex. att Skatteverket går ut och informerar om en stor kontrollinsats mot en viss bransch eller företeelse. Mycket talar för att fler i den utpekade målgruppen upplever att risken för upptäckt ökar, men hur reagerar övriga skattskyldiga? Tänker företag i andra branscher att risken för upptäckt minskar för dem eftersom Skatteverket ändå är upptagna på annat håll? Hur ser man på tidsaspekten? Tänker de branscher som blir granskade att risken för upptäckt ökar endast under kontrollperioden, och minskar efteråt, eller ser de en permanent ökad risk för upptäckt? Hur påverkas den upplevda risken för upptäckt om Skatteverket går ut och kontrollerar en viss företeelse men förbiser andra fel? Minskar den upplevda risken för upptäckt för denna andra typ av fel?

Länk 3 Beteendeförändring

Även länken mellan upplevd risk för upptäckt och påverkan på skattefelet är osäker. Kedjan bygger på antagandet att kontrollinsatsen medför att den upplevda risken för upptäckt ökar och att detta påverkar skattefelet. Är då detta alltid fallet? Mycket talar för att olika skattskyldiga reagerar på olika sätt på samma upplevda risk för upptäckt. Det finns andra faktorer än upplevd risk för upptäckt som påverkar beteendet. En ekonomiskt och socialt framstående person har i regel mer att förlora på att framstå som en fuskare eller slarver än en person som redan åtskilliga gånger varit föremål för olika myndigheters åtgärder och kontroller. Exempelvis har antagligen en trogen kund hos Kronofogden inte mycket att förlora på marginalen om ytterligare några tusenlappar restförs, om där redan finns stora belopp.

Hela effektkedjan

När det gäller den allmänpreventiva effekten finns betydligt fler osäkerheter än det gör för den individualpreventiva effekten. Den första stora osäkerhetsfaktorn är att den information som är tänkt att påverka de skattskyldiga förmedlas via medier. En faktor som inte kan kontrolleras av Skatteverket. Beroende på hur denna information ser ut kan påverkan bli en helt annan än vad som är avsikten från skatteverkets sida. Även den skattskyldiges riskbedömning är svårförutsägbar. Beroende på hur den individuella risken för upptäckt bedöms kan effekten bli den rakt motsatta mot den önskvärda. Även om den upplevda risken för upptäckt är identisk mellan två personer kan beteendet vara olika beroende på vilka effekter en upptäckt skulle få. Till sist är det ur avgränsningssynpunkt mycket besvärligt att använda skattefelet som slutlig effekt. Sammanfattningsvis så finns ännu större utforskade områden vad gäller den allmänpreventiva effekten än vad som finns om den individualpreventiva effekten.

Sammanfattning

När alla potentiella effekter beaktas går det inte att fastslå att det finns ett entydigt samband mellan kontroll och skattefel, på så sätt att en kontrollinsats alltid leder till minskat skattefel. Hur enskilda insatsers effekter påverkar skattefelet är en empirisk fråga.

6.2 Målet och dess mätbarhet

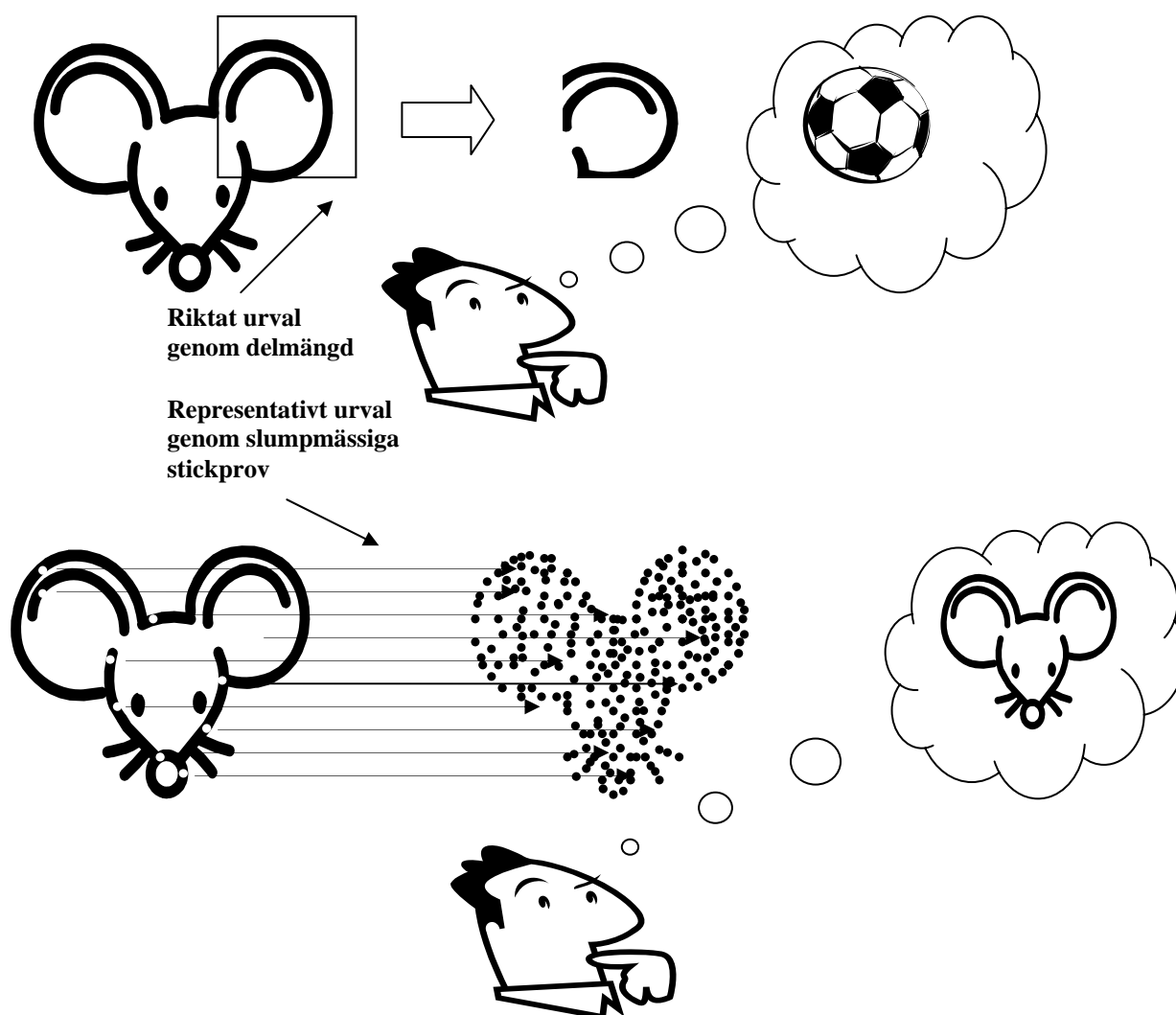
Det har tidigare konstaterats att det övergripande målet inte är direkt mätbart, vare sig i sig självt eller sett över tiden som en förändring. I detta avsnitt diskuteras huruvida det finns sätt att indirekt mäta eller uppskatta olika insatsers påverkan på det övergripande målet. Det handlar alltså om att hitta alternativa målvariabler som ersätter en direkt mätning av påverkan på det övergripande målet, samtidigt som olika svårigheter beaktas.

6.2.1 Validitet och reliabilitet

I resonemanget nedan kommer begreppen validitet och reliabilitet att användas. Validiteten handlar om att man mäter det man vill mäta, vad som mäts, reliabiliteten om att mätningen är pålitlig, hur det mäts. Ett enkelt exempel beskriver begreppen. Antag att man vill minska skattefusket i en viss bransch genom en informationskampanj. Om man skulle bestämma sig för att använda antalet påförda förseningsavgifter som mått på kampanjens framgång så skulle det inte vara särskilt lyckat. Antalet förseningsavgifter säger inte så mycket om skattefusket totalt sett. En undersökning med indikatorn förseningsavgift skulle med andra ord ha låg validitet. Om man istället skulle ta en variabel med högre validitet, t.ex. hur många rättelser av deklarationerna som görs före och efter informationskampanjen, så finns andra osäkerhetsfaktorer. Mellan mätpunkterna kan många andra faktorer än informationskampanjen ha påverkat skattebetalarna. Det går inte att säkert säga att just informationskampanjen är orsak till en eventuell förändring. Om inte hänsyn tas till denna osäkerhet har mätningen låg reliabilitet.

En annan fråga som har med validiteten att göra är representativiteten. I många sammanhang vill man uttala sig om en stor grupp som det inte finns några praktiska möjligheter att undersöka i sin helhet. Det kan handla om skattebetalare som kännetecknas av en viss företeelse, en viss bransch eller hela skattebetalarkollektivet. I sådana sammanhang måste man göra ett urval som är representativt för en större population. Det vanligaste sättet att göra detta är att göra ett slumpmässigt urval. Om urvalet däremot gått till på något annat sätt, som gör att det inte är representativt för hela populationen, går det inte att uttala sig om andra än just de som blivit granskade. Ett exempel på en annan typ av urval som är mycket vanligt i samband med Skatteverkets granskningsinsatser är när inriktningen är begränsad till dem som anses mest granskningsvärda. Det handlar alltså om ett riktat eller spetsigt urval. I sådana fall handlar det om totalundersökningar av särskilda delmängder av populationen. Att från dessa dra slutsatser om hela populationen kan vara mycket missvisande. Hur det ser ut bland alla övriga utanför delmängden är och förblir okänt.

Fig 6.5 Riktat urval och representativt urval



6.2.2 Mätbarhet genom operationalisering av effektmålet

I de allra flesta fall går det inte att direkt mäta effekten på det övergripande målet. Detta gäller, som redan konstaterats flera gånger, i högsta grad olika insatsers påverkan på skattefelet. Nedan listas ett antal praktiska och teoretiska svårigheter med att använda skattefelets förändring som mål.

I. Mycket hypotetiskt

Skattefelet, definierat som differensen mellan den skatt som skulle betalas om alla skattepliktiga aktiviteter redovisades korrekt och den skatt som faktiskt betalas, är mycket hypotetiskt. Detta eftersom många skattepliktiga aktiviteter aldrig hade ägt rum om de skulle ha beskattats. Istället skulle andra aktiviteter ha ägt rum. En tänkt heltäckande beskattning slår alltså ut vissa aktiviteter medan nya till viss del skulle tillkomma. En allomfattande beskattning skulle alltså utlösa en kedja av effekter som är omöjliga att överblicka.

II. Ej observerbart

Långtifrån alla skattepliktiga aktiviteter är observerbara. Större delen av de obeskattade aktiviteterna företas i det fördolda. Lika dolda som de är för Skatteverkets kontroller är de för mätningar.

III. Oklara gränser

Det finns en zon där oklarhet råder om skatteplikt råder eller ej. Ett exempel är tjänster som utförs av och för privatpersoner. Är det skattepliktigt om:

- Ett minderårigt barn dammsuger hemma och får en peng av mamma?
- Om ett barn i gymnasieåldern regelbundet klipper gräs i hemmet?
- Om samma barn samtidigt klipper gräs åt grannen?
- Om rörmokaren hjälper sin granne revisorn med målning samtidigt som revisorn hjälper rörmokaren med att ansöka om bygglov?
- Om samma rörmokare hjälper revisorn med ett läckande rör samtidigt som revisorn hjälper rörmokaren med bokslutet till dennes rörelse?

Någonstans i denna uppräkningsgång går en gräns för när beskattning bör ske, men var?

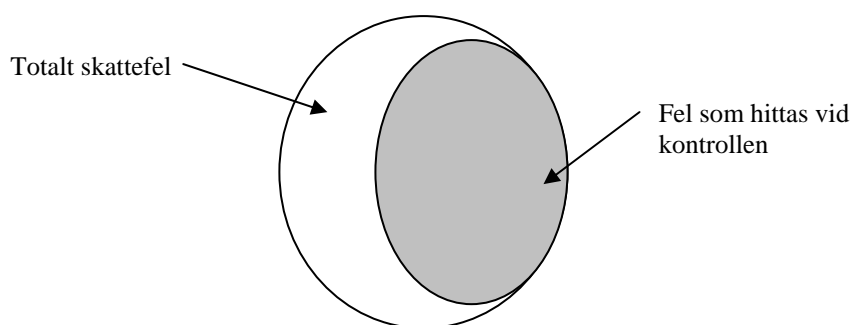
IV. Skatteeffekten

Även om alla skattepliktiga aktiviteter vore observerbara kan man endast mycket ungefärligt beräkna storleken på de skatter som skulle ha belöpt på dem. En viktig orsak till detta är att samma typ av skattepliktiga inkomster kan ha olika skatteeffekt på grund av stigande marginalskatt. En annan orsak är att samma underlag kan belastas med många olika skatter och avgifter, tex inkomstskatt, moms och arbetsgivaravgifter.

Det kan alltså konstateras att skattefelet är en hypotetisk konstruktion som inte direkt låter sig mätas. Om skattefelet och dess förändring trots de praktiska och teoretiska svårigheterna ska användas som målvariabel måste en operationell definition användas. En operationell definition är en bestämning som är användbar i praktiken. En tänkbar operationell definition av skattefelet i samband med en kontrollinsats är att låta det vara detsamma som alla fel som hittas vid kontrollen, dvs alla fel som identifieras anses också vara alla existerande fel. De oredovisade underlag som sedan hittas får sedan multipliceras med en fiktiv skattesats.

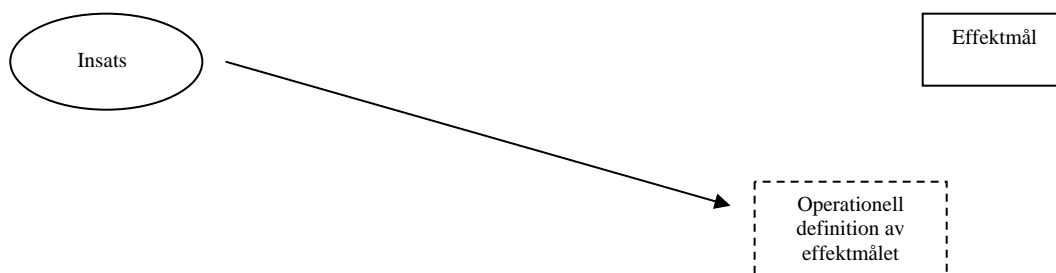
När man använder en operationell definition gäller det dock att vara medveten om validiteten. Det mest uppenbara validitetsproblemet i detta sammanhang är den eventuella bristen på överensstämmelse mellan det man vill mäta, dvs det teoretiska skattefelet, och det man kan mäta, dvs den operationella definitionen av skattefelet.

Fig 6.6 Det verkliga skattefelet och en operationell definition av det



Hur noggranna kontroller som än görs kommer alla fel som utgör differensen mellan den teoretiskt korrekta skatten och den faktiska skatten aldrig att påträffas. Därför är det totala skattefelet större än de fel som hittas vid kontrollen. Hur stor skillnaden är kommer vi dock aldrig att kunna ta reda på. Detta är en osäkerhet som måste beaktas om en operationell definition används för att mäta en insats effekter på det övergripande målet. Det kan vara så att en operationell definition av det övergripande målet inte leder ända fram. Risken finns också att det rentav leder insatserna fel.

Fig 6.7 En operationell definition av effektmålet kan leda fel

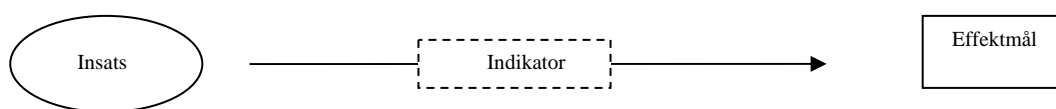


Ett annat problem som kan leda både till över- och underskattningar av skattefelet är om resultatet av en kontrollinsats mot en mindre grupp får representera skattefelet för en större grupp. Detta är ett representativitetsproblem som är särskilt vanligt när kontrollerna är inriktade mot dem som anses mest kontrollvärda. I sådana fall torde det vara vanligt med överskattningar av det totala skattefelet.

6.2.3 Mätbarhet genom indikatorer som ligger i linje med effektmålet

Enligt resonemanget i avsnitt 6.1.2 går det inte att utan vidare utgå från att en kontrollinsats alltid leder till önskad effekt på det övergripande målet, definierat som skattefelet. Enligt resonemanget ovan finns det också problem med att direkt mäta effekterna på skattefelet. Frågan är då om det finns andra metoder att mäta olika insatsers effekter på det övergripande målet? Ett vanligt sätt att angripa problemet är att ta fram mätbara indikatorer som anses relaterade till effektmålet. Det hela kan illustreras som nedan.

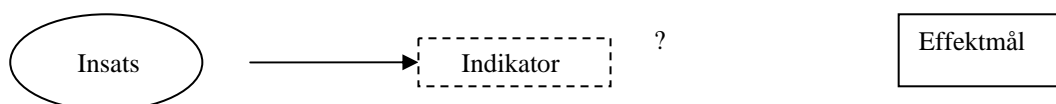
Fig 6.8 En lämpligt indikator bör ligga i linje med effektmålet



Det första att beakta i sammanhanget är att indikatorn verkligen ligger i linje med effektmålet. Att hitta relevanta indikatorer är dock svårt. Det är därför ofta frestande att använda de prestationer som utförs som indikatorer. De är lätta att iaktta, räkna och redovisa. När det gäller prestationer är det dock sällan som de självklart leder till den önskade effekten.

Följande exempel kan beskriva problemet. Antag att man genom en brevkampanj vill minska skattefelet i en viss bransch. Om antalet utskickade brev skulle utgöra indikator, och effekten enbart skulle mätas genom detta mått, skulle det inte gå att säga något om hur framgångsrik kampanjen varit. Det finns inget självklart samband mellan utskickade brev och ändrat beteende.

Fig 6.9 En indikator säger inte alltid något om uppfyllelsen av effektmålen



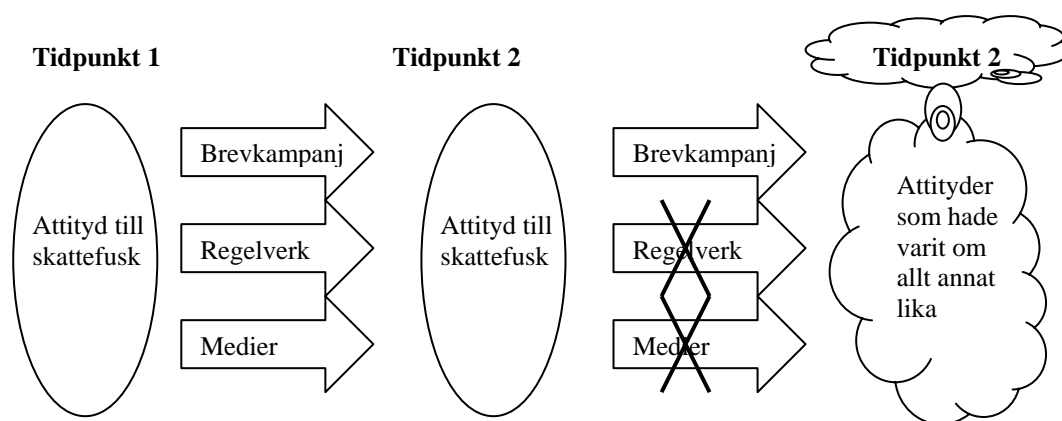
En mer relevant indikator i sammanhanget skulle kunna vara en förändrad attityd gentemot skattefusk, ökat antal självrättelser i branschen eller något annat mål närmare det formulerade effektmålet.

Tyvärr kan det dock uppstå andra problem med indikatorer som är relevanta. Ett problem med många relevanta indikatorer är att de är svåra eller dyra att mäta. Avvägningen mellan relevans och mätbarhet å ena sidan och svårigheter och kostnader å den andra måste bedömas från fall till fall. Denna problematik behandlas ytterligare i del 3. Ett annat problem är att indikatorn i sig inte alltid säger tillräckligt om önskvärda effekter, även om den är relevant. Det kan nämligen finnas andra faktorer som också kan inverka på den eftersträlvade effekten. Problemet med så kallade externa faktorer beskrivs närmare nedan.

6.2.4 Externa faktorer och reliabiliteten

En vanlig ansats är att mäta hur det ser ut före en insats för att sedan mäta hur det ser ut efter den. Kan man med detta vara helt säker på att skillnaden mellan före och efter beror på insatsen? Det kan man i regel inte. En viktig orsak till att man inte kan det är andra påverkansfaktorer, dvs förutom insatsen kan det under den tiden som gått mellan för- och eftermätning ha inträffat andra saker som också haft en påverkan. I exemplet ovan där attityden till skattefusk nämns som exempel på indikator är det högst angeläget att beakta andra påverkansfaktorer. I detta fall kan sådana faktorer som förändringar i regelverket och medierapportering vara värda att beakta. Exempelvis kan ändringar i regelverket, som innebär en bättre anpassning till allmänhetens uppfattning om vad som är rimligt och rättvist, ha påverkat attityderna mellan mätningarna. Ett annat exempel på faktorer som kan ha påverkat är medierapportering om offentliga personers fusk.

Fig 6.10 Externa påverkansfaktorer



Det beskrivna problemet med andra faktorer handlar om reliabilitet, dvs. om mätningens pålitlighet. Ett sätt att hantera problemet är att försöka isolera brevkampanjens effekter från effekterna av andra faktorer, genom att ha en referensgrupp som inte skiljer sig från den kontrollerade gruppen på annat sätt än att den inte blivit föremål för kampanjen. Denna och andra metoder för att isolera effekter behandlas utförligare i del 3.

6.2.5 Interna faktorer och målförskjutning

Indikatorer som inte ligger i linje med de önskade effekterna kan bli styrande. Det handlar här om problemen med målförskjutning som även beskrivits i avsnitt 3.8.2.

Målförskjutning kan exempelvis inträffa när prestationer ensamma används som indikatorer på effektmålsuppfyllelse. Ett exempel på när det finns risk för målförskjutning är om höjningsbelopp används som enda kriterium för framgång i kontrollverksamheten. Att ha höjningsbelopp som enda kriterium innebär att de ska maximeras. Detta innebär att ju fler skattebetalare som gör fel desto bättre möjligheter finns att åstadkomma ett bra resultat - enligt indikatorn.

Det kan här vara på sin plats att understryka att resonemanget handlar om effektmätning och att det i andra sammanhang mycket väl kan vara motiverat att mäta prestationer. T.ex. är produktivitet också viktiga aspekter av verksamheten.

Olämpliga indikatorer kan alltså leda till att resultatet blir något helt annat än det önskvärda. Risken för detta är särskilt stor när det man vill mäta är svårt att mäta. Man mäter istället då gärna det som går att mäta. Innan man sätter igång med att ta fram indikatorer måste man därför se till att man mäter det man vill mäta, dvs att indikatorn ligger i linje med effektmålet. Vilka indikatorer som ligger i linje med effektmålet går inte att säga generellt. Det måste bedömas från fall till fall. När det gäller kontrollinsatser skulle dock en tänkbar indikator vara att undersöka hur beteendet hos ett representativt urval skattskyldiga påverkas av kontrollinsatsen.

6.3 Sammanfattning och slutsats kapitel 6

Det har flera gånger i detta avsnitt konstaterats att det inte finns någon möjlighet att mäta enskilda insatserns påverkan på de långsiktiga målen. Det bästa som kan göras istället är att ta fram indikatorer på måluppfyllelse. Det finns dock många omständigheter att beakta när sådana ska tas fram. Ovan har några av de viktigaste beskrivits.

- Indikatorn ska vara relevant men inte för dyr eller komplicerad att mäta,
- Indikatorns effekter ska kunna isoleras från effekterna från andra faktorer
- Indikatorn får inte vara av sådant slag att resurserna styrs fel eller används på ett inte optimalt sätt

Dessvärre finns ingen standardmodell som kan användas för att ta fram och mäta indikatorer. Hur en effektmätning ska genomföras i praktiken får istället avgöras från fall till fall. Vilka metoder som kan användas, avvägningar som ska göras och möjligheter som finns är ämnet för nästa avsnitt, del 3 *Metoden för effektmätning*.

Del III: Metoden för effektmätning

7 Inledning: Effekter och modeller för att beskriva insatsen

Denna del är mer praktiskt inriktad och behandlar rent metodmässiga och hantverksmässiga aspekter av effektmätning. Syftet är att på ett överskådligt sätt behandla effektmätning för det slags insatser som Skatteverket genomför.

I detta inledande kapitel 7 tittar vi närmare på vad som menas med effekter och ett par modeller som kan användas för att åskådliggöra vilka effekter en insats kan tänkas ge och varför. I kapitel 8 diskuterar vi grunderna i effektmätning genom att betrakta tre kvalitetskriterier för en effektmätning. För varje kvalitetskriterium presenteras grundläggande teori och praktiska råd. Kapitel 9 studerar de särskilda förutsättningar som gäller för effektmätning av Skatteverkets insatser. Kvalitativa metoder för att beskriva effekter diskuteras i kapitel 10. I det avslutande kapitel 11 diskuteras när det är meningsfullt att försöka mäta effekterna av en insats, när extra resurser för mätning kan vara motiverat samt hur resultatet av effektmätningen ska tas om hand.

I sammanfattningen presenteras ett praktiskt hjälpmedel i form av ett flödesschema för effektmätning. Flödesschemat, som är utformat för att användas då den insats som eventuellt ska effektmätas befinner sig på planeringsstadiet, avser att vägleda såväl beställaren av insatsen som utföraren av effektmätningen genom effektmätningens olika steg och beslutpunkter. Att ha flödesschemat som utgångspunkt vid läsning av rapporten underlättar förmodligen överblicken. Observera att flödesschemat följer en delvis annan disposition än rapporten.

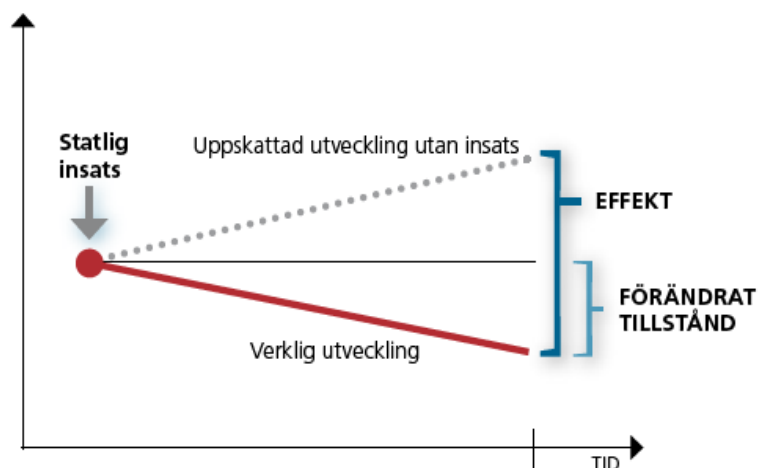
7.1 Effekter

Skatteverket agerar i en komplex omvärld i ständig förändring, och ämnar genom sina insatser att förändra något så komplext som mänskligt beteende – allt fler ska göra rätt för sig. Det finns därför ett stort behov av att försöka isolera effekterna av själva insatsen som ska effektmätas från effekterna av andra påverkansfaktorer, t.ex. konjunkturen.

För att kunna säga vilken effekt själva insatsen har gett måste vi alltså kunna säga hur tillståndet skulle ha sett ut om insatsen inte hade genomförts. Detta benämns som det *kontrafaktiska tillståndet*. Den eventuella effekten motsvaras av skillnaden mellan det faktiska och det kontrafaktiska tillståndet. Idealt sett isoleras effekten av själva insatsen från effekten av andra tänkbara påverkansfaktorer. Observera att effekten kan vara större än själva förändringen (se figur 7.1).³¹

³¹ Figur från ESV 2006:8.

Figur 7.1 Illustration av begreppen effekt och kontrafaktiskt tillstånd



Så långt det är möjligt vill vi göra en god bedömning av det kontrafaktiska tillståndet. Men att göra detta, och således isolera effekterna av en insats, innebär i allmänhet en stor utmaning. I det kommande ska vi se vad som krävs för att detta ska vara möjligt. I korthet handlar det om att utfallet, utöver insatsen, endast i begränsad omfattning påverkas av föränderliga faktorer, att det "naturligt" finns referensgrupper som väl beskriver det kontrafaktiska tillståndet, eller att insatsen utformas på ett sådant sätt att effekten av insatsen kan isoleras.

7.2 Verksamhetslogik som hjälpmedel vid effektmätningar

I avsnitt 3.1 beskrev vi hur man skissa på en verksamhetslogik för att bl.a. belysa vilka eventuella effekter en insats har gett upphov till och varför. Vi rekommenderar att verksamhetslogiken används som analytiskt verktyg och utgångspunkt i effektmätningen. Dessutom kan resultatet av genomförd effektmätning tolkas utifrån verksamhetslogiken (se 11.3).

Bärande i verksamhetslogiken är den s.k. effektkedjan, men den behöver kompletteras med beskrivningar av s.k. ramfaktorer och externa faktorer. Dessa tre komponenter utgör tillsammans den utbyggda verksamhetslogiken och i detta avsnitt diskuteras hur denna kan användas som praktiska verktyg i genomförandet av effektmätningar.

En *effektkedja* beskriver hela den serie av orsak-verkan-samband som förenar den aktuella insatsen med den tänkta effekten. En sådan kedja är begränsad till just de faktorer som binder samman insatsen med effekten. I den enklaste av världar där insatsen är både nödvändig och tillräcklig för effekten ger effektkedjan en totalbild av förloppet. I verkligheten finns dock även andra faktorer som kan påverka bilden.

Ramfaktorer är sådana faktorer som gör att samma åtgärd kan leda till ett annat resultat i en annan kontext, miljö, situation etc. Ramfaktorer handlar om sådant som kultur, traditioner, arbetsklimat, etc.

Externa faktorer är sådana faktorer som ligger utanför effektkedjan, d.v.s. i omvärlden, som också kan påverka utfallet. Exempel är attitydförändringar, konjunkturförändringar, mass-mediala händelser, etc.

Observera skillnaden mellan ramfaktorer, som är möjliggörare eller hinder/risker för att insatsen ska uppnå målen, och externa faktorer, som gör det svårt att isolera effekten av själva insatsen. De förra kan med fördel betraktas som en del av insatsen, eller åtminstone dess naturliga kontext, medan de senare "endast" försvårar mätningen.

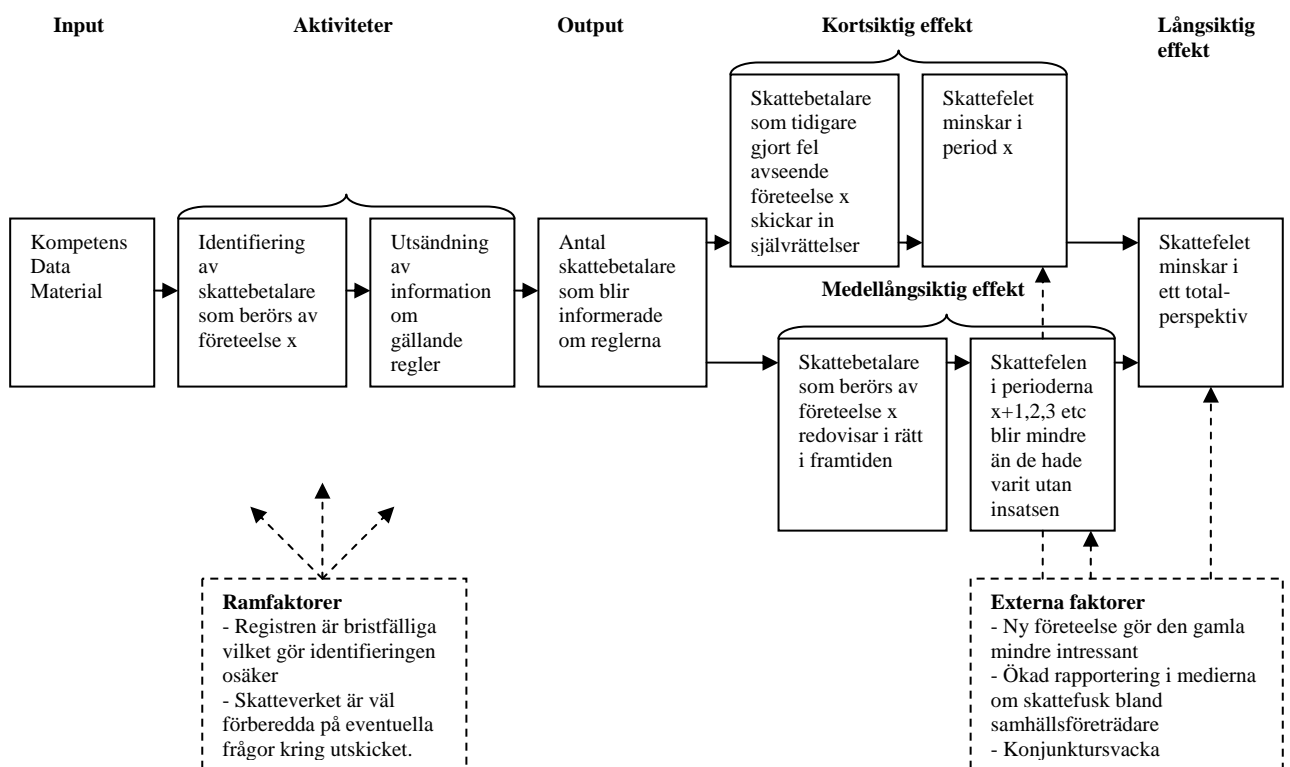
Exempel på beskrivning av en insats med hjälp av verksamhetslogik

Nedan beskrivs hur en tänkt insats kan beskrivas med hjälp av en uppställd verksamhetslogik.

Antag att Skatteverket planerar att genomföra en riktad informationskampanj mot skattebetalare som berörs av en företeelse som man vet är förenad med en stor mängd fel. Kampanjen är tänkt att gå till på så sätt att man genom information i Skatteverkets interna register identifierar dem som berörs av företeelsen. Till dessa skickas därefter information om vilka regler som gäller avseende företeelsen. Tanken är att skattebetalarna som gjort fel därefter dels ska skicka in rättelser av redan lämnade deklarationer, dels ska se till att deklarationer som lämnas i framtiden är riktiga i det aktuella avseendet. Skattefelet avseende den period som rättas minskar därmed som en följd av rättelsen. Skattefelet kommande perioder blir också lägre än de annars skulle ha varit. Därmed blir också skattefelet sett i ett totalperspektiv mindre. Observera att effekterna bör betraktas i ett tidsperspektiv.

Till effektkedjan ska läggas ramfaktorer och externa faktorer. Det går att tänka sig en mängd sådana. I de prickade boxarna nedan ges exempel på några. Vi återkommer strax till detta exempel för att se hur verksamhetslogiken kan fungera som analytiskt verktyg.

Figur 7.2 Illustration av en effektkedja samt ramfaktorer och externa faktorer



7.2.1 Hur kan verksamhetslogiken användas i praktiken?

Den praktiska nyttan med att ha identifierat hela den utbyggda verksamhetslogiken är flerfaldig:

1. Den övergripande nyttan uppstår då en effektmätning eller utvärdering ska planeras, diskuteras och kommuniceras. De händelseförlopp som ska iakttas, mätas och bedömas är i regel komplexa och svåröverskådliga. Att rita upp en effektkedja med alla tänkbara externa

faktorer och ramfaktorer är därför ett mycket bra sätt att få förloppet klart för sig. Den är också en mycket bra utgångspunkt när flera personer är inblandade i uppgiften. Detta arbetssätt kan bidra till att skapa en gemensam syn på insatsen.

2. Effektkedjan illustrerar *varför* insatsen förväntas leda till effekt. Frågan är om den väl fångar det tänkta förloppet, eller behöver kompletteras för att fånga händelseförloppets komplexitet.

3. Den uppställda kedjan illustrerar hur och hos vem/vilka effekterna är tänkta att uppstå. Den eller de långsiktiga effekterna är dock ofta inte direkt mätbara, utan vi kan behöva identifiera lämpliga indikatorer på dessa, eller begränsa effektmätningen till att omfatta en eller flera deffekter. Kedjan illustrerar då vilka antaganden som kopplingen mellan deffekten och den slutliga effekten innebär. Utifrån detta gör vi en bedömning av vilka länkar som kan fastställas a priori (utan hänvisning till annan erfarenhet) och vilka länkar som behöver testas *empiriskt* om de håller.

4. Genom att synliggöra eventuella föränderliga externa faktorer som också kan påverka utfallet kan vi göra en bedömning av om det krävs ett särskilt upplägg på insatsen för att isolera dess effekter från dessa externa faktorer (se vidare avsnitt 8.2). Observera att det i normalfallet är svårare att isolera effekterna av insatsen, ju längre tid som förlöper efter insatsen.

5. Genom att identifiera de ramfaktorer som är särskilt betydelsefulla för insatsens effekter, kan vi försöka stärka dessa och bidra till att effekterna av insatsen blir större.

Exempel på hur verksamhetslogiken kan användas i praktiken

Nedan analyserar vi det exempel vi beskrev ovan.

Kedjans hållbarhet

Kedjan ovan är mycket enkel. För att ordentligt undersöka hela förloppet, med alla mellanliggande händelser och tänkbara alternativa förlopp, skulle krävas en längre kedja med en mängd förgreningar. I denna mening är kedjan ofullständig. Den avser dock att teckna en sammanfattande bild om hur och varför insatsen förväntas leda till de tänkta effekterna, och ger därför en något idealiserad bild av insatsen.

Om man vill komplettera effektkedjan kan man t.ex. tänka sig ytterligare ett flertal steg mellan den tredje rutan, utsändning av information, och de efterföljande. En ytterligare ruta skulle kunna vara att skattebetalarna får del av informationen. Det är alls inte säkert att de får det bara för att informationen skickas ut. Det skulle kunna vara så att brevet som den sänds ut i är mycket likt ett reklambrev som råkar bli skickat samtidigt och att många därför slänger det oöppnat. En ytterligare ruta skulle kunna vara att skattebetalaren förstår informationen, en annan att denne får eventuella frågor kring utskicket pedagogiskt besvarade av Skatteverket.

Sammanfattningsvis kan vi inte vara säkra på att kedjan i exemplet håller. Detta behöver testas empiriskt.

Identifierade deffekter och indikatorer

Det övergripande målet ter sig mycket svårfångat (som vi sett i kapitel 6). Effektkedjan pekar dock på möjligheten att vi begränsar mätningen till att omfatta delmål, t.ex. att fler gör själv-rättelser och att fler gör rätt kommande perioder. (Vi återkommer till detta i avsnitt 8.1.)

Den kan också illustrera under vilka antaganden som kopplingarna mellan de korta och de långsiktiga effekterna görs. Antagandena bygger på att informationsinsatsen bidrar till att fler gör rätt innevarande och kommande taxeringsperioder, och att skattebetalarna inte – åtminstone inte i motsvarande omfattning – gör *mer* fel inom andra områden, som följd av ökat fokus från Skatteverkets sida på detta område.

Externa faktorer

Vi pekade ovan på möjligheten att effektmätningen ska omfatta endast de effekterna att fler gör självrättelser och mer rätt även i framtiden. Vi behöver identifiera externa faktorer som kan påverka utfallet, d.v.s. uppfyllelsen av dessa effektmål. Man kan t.ex. tänka sig att denna informationsinsats sammanfaller med en massmedial debatt om fel och fusk eller att företeelsen dramatiskt minskar i omfattning till nästkommande år p.g.a. en konjunktursvacka. Det kan också vara så att informationsinsatsen är den enda större händelsen kopplad till denna företeelse. Ju längre tid som fortlöper efter informationsinsatsen, desto svårare är det förstås att isolera effekten av själva insatsen.

Ramfaktorer

Som tidigare nämnts är ramfaktorer sådant som gör att en och samma åtgärd kan ge ett annat resultat i ett annat sammanhang. Det går att tänka sig en mängd sådana faktorer i exemplet ovan. En sådan är om registren som används för adressering av informationen är så bristfälliga att många av dem som får den överhuvudtaget inte är berörda, samtidigt som många som faktiskt är det inte får informationen. Insatsen skulle i sådant fall få en sämre effekt än om registren hade innehållit mer relevant information.

Den utökade programteorin ger oss en god grund utifrån vilken vi empiriskt kan studera effekterna av insatsen. Det första steget i flödesschemat för effektmätning är därmed avklarat.

8. Att mäta effekter – kvalitetskriterier för en effektmätning

I detta kapitel kommer vi att uppehålla oss vid vilka möjligheter och begränsningar som gäller generellt då vi kvantitativt vill bedöma effekternas storlek, d.v.s. att *mäta* effekterna. I nästa kapitel kommer vi att se vilka särskilda förutsättningar som gäller för effektmätning av den typ av insatser som Skatteverket genomför. I kapitel 10 kommer vi att vidga perspektivet och se hur effekterna kan belysas med hjälp av kvalitativ metod.

När det handlar om kvaliteten på effektmätningen ser vi att det är tre kriterier som är särskilt centrala:

- Kvaliteten på effektmåtten. Detta kriterium rymmer två frågor:
 - I vilken mån återspeglar effektmåtten de verkliga effekterna, och
 - i vilken mån är mätningen pålitlig?
- Isolerbarhet: I vilken mån är den eventuellt observerade förändringen av effektmåtten över tiden (före och efter insatsen) ett resultat av själva insatsen?³²
- Generaliserbarhet: För vilka är resultatet giltigt?

Kvaliteten på effektmåtten och isolerbarheten är gemensamt de kriterier som avgör i vilken mån det är möjligt att teckna en god bild av effekterna av en insats. Tillgången på relevant data är central. För att mäta det man vill, med god kvalitet, behöver man ofta *före* insatsen

³² Formuleringen är medvetet något förenklad. Som figur 7.1 illustrerade är det skillnaden mellan det faktiska och kontrafaktiska tillståndet som definierar effekten. Ett oförändrat värde på effektmåttet kan således också innebära en effekt, på samma sätt som effekten kan vara större än den observerade förändringen.

påbörjas skapa förutsättningar för relevant datainsamling. Val av utformning och upplägg av insatsen påverkar möjligheten att isolera effekterna av själva insatsen, men också *generaliserbarheten* – för vilka resultatet av effektmätningen blir giltigt.

Ovanstående illustrerar att det är av stor vikt att redan på insatsens planeringsstadium fundera över hur insatsens effekter ska mätas.

I de kommande avsnitten kommer vi att studera vart och ett av de tre kvalitetskriterierna i mer detalj.

8.1 Kvaliteten på effektmåtten

8.1.1 Bestäm utifrån effektkedjan vad mätningen ska omfatta

I avsnitt 7.2 diskuterade vi effektkedjan som ett verktyg för att beskriva och utvärdera eventuella effekter av en insats. De länkar i kedjan som inte kan antas giltiga a priori, måste testas empiriskt. Det är sällsynt att det går att direkt mäta i vilken grad insatsen gav upphov till den långsiktiga effekten i kedjan. Oftast behöver vi antingen avgränsa mätningen till att omfatta en *deleffekt* som ligger i linje med den långsiktiga effekten, t.ex. en kortsiktig effekt, eller försöka formulera en indikator på den långsiktiga effekten. Utifrån effektkedjan bestäms vilka effektmått som mätningen ska omfatta.

Ur effektmätningsspektiv kan det även, i vissa fall, vara värdefullt att mäta enbart prestationerna. Om det t.ex. redan i effektkedjans prestationsdel finns länkar som riskerar vara svaga, kan en mätning av dessa ge stöd för att insatsen har gett endast små effekter.³³ Ett annat exempel är då det finns mycket goda skäl att förvänta sig att en prestation leder till önskad effekt, t.ex. att en betydande växling av färdmedel för tjänsteresor från flyg till tåg leder till en bättre miljö, eftersom koldioxidutsläppen minskar.

Det handlar alltså om att bestämma såväl *vad* som ska mätas och *när*. De direkta effekterna kan förväntas uppträda – och avta – snabbare än t.ex. en indikator på den långsiktiga effekten. Om möjligt bör mätningen omfatta såväl indikatorer på korta, medellånga och långa effekter.

Innan vi går vidare och talar om teoretiska och operationella definitioner av dessa effektmått ska vi orientera oss i vilka metoder för datainsamling som kan användas.

8.1.2 Olika metoder för datainsamling

Skatteverket har en rik mängd data. För effektmätningar är det informationen om skattebetalarna som är central. Det kan vara information som skattebetalaren själv tillhandahåller (genom deklarationsblanketter, registreringsanmälningar, m.m.), information från tredje part (Bolagsverket, Vägverket, etc.), information om när skattebetalaren t.ex. har deklarerat eller satt in pengar på Skattekontot och sådan information som Skatteverket aktivt inhämtar genom t.ex. en kontrollaktivitet (beslut efter utredning, kontrollmaterial, etc.).

För den praktiska informationsinsamlingen är det viktigt att fundera över vilken data som redan finns och löpande görs tillgänglig (i Skatteverkets databaser, offentlig statistik,

³³ Observera att motsatsen inte kan antas utan vidare, d.v.s. goda prestationer ger inte med nödvändighet goda effekter, men dåliga prestationer kan inte leda till goda effekter. Låt oss knyta an till exemplet i föregående kapitel och anta att identifieringen av vilka skattebetalare som omfattas av företeelse x är bristfällig och att informationsinsatsen riktas fel. Effekten av insatsen blir då, på sin höjd, liten.

diarieförda handlingar, etc.) och vilken information som kan tillgängliggöras (via t.ex. enkäter, intervjuer och kontroller).

Vi behöver fundera över vilken av all potentiellt tillgänglig data som kan associeras med de effekter som effektmätningen ska omfatta. Ledstjärnan är att man använder sig av *flera* metoder för datainsamling, s.k. triangulering. Det ger en mer nyanserad och mångfacetterad bild av det man studerar. Även om den potentiellt tillgängliga informationen sätter ramarna bör perspektivet vara allt annat än snävt, snarare kreativt och nytänkande. Att identifiera relevanta källor eller metoder för informationsinhämtning, även om dessa inte tidigare har använts, är mycket viktigt.

Viktigt är också att i god tid planera för informationsinhämtning och att informationen som inhämtas under en insats lagras per skattebetalare, och inte endast på en aggregerad nivå.³⁴ Detta är många gånger helt nödvändigt för att effekterna av insatsen ska kunna uppskattas överhuvudtaget, samt för att göra det möjligt att fördjupa analysen av insatsens effekter (t.ex. att studera hos vilka grupper av skattebetalare som den eventuella effekten är som störst).

8.1.3 Teoretisk definition

Vi behöver först definiera de begrepp som effektmätningen ska omfatta. I denna fas ska vi *inte* bekymra oss om definitionen underlättar eller försvårar mätningen. Definitionen ska vara sådan att den på bästa sätt klargör innebörden i begreppet.

I SKV 2006:3 *Krånglig moms – en företagsbroms?* återges Cedric Sanfords definition av fullgörandekostnader som de kostnader som uppstår i samband med att skattebetalarna fullgör sina skyldigheter enligt gällande skattelagstiftning och skatteadministrativa rutiner. Exempel på andra begrepp som kan behöva definieras är betalningsbeteende, förtroende och oavsiktliga fel. Det är på faktorer som dessa som vi önskar att våra insatser har effekt.

Observera att många relevanta begrepp har definierats inom ramen för tidigare insatser eller strategiska rapporter. Är begreppen väletablerade och lämpliga – använd då dessa.

8.1.4 Operationella definitioner: effektmått

Vi behöver gå från en teoretisk definition av begreppet, till att hitta en s.k. operationell definition. Den operationella definitionen av begreppet beskriver hur man tänker sig att mäta det.

Den potentiellt tillgängliga informationen (se 8.1.2) sätter ramarna för vad som är möjligt att beskriva och mäta. Utifrån denna information kan operationella definitioner formuleras. Låt oss benämna dem för effektmått. De kan vara en eller flera till antalet.

Kvalitetskriterier

Två huvudsakliga kvalitetskriterier finns för effektmåtten.

- Målet är att mäta rätt saker...

Den operationella definitionen fångar väl den teoretiska. Eller, för att uttrycka det med andra ord, effektmåttet beskriver väl den effekt som insatsen syftar till.

- ...på ett tillförlitligt sätt.

³⁴ Givetvis måste behandlingen av personuppgifter följa bestämmelserna i registerförfattningarna och personuppgiftslagen.

Förutom en god kvalitet på den operationella definitionen är själva mätningen av de mått vi preciserat av största vikt.

Ett par exempel:

- Vid mätning av effekterna av införandet av personalliggarna i restaurang- och frisörbranschen är antalet AG-registreringar och redovisat AG-underlag exempel på effektmått som svarar väl mot de effekter av minskat svartarbete man önskar uppnå med insatsen. Tillförlitligheten i mätningen är perfekt i den meningen att uppgifterna är de som skattebetalarna har lämnat.
- Utfallet av en skrivbordskontroll är ett mycket osäkert mått på det eventuella totala skattefelet hos en företagare. Utfallet av en omfattande revision har dock potential att väl fånga det eventuella skattefelet. Tillförlitligheten i själva ”mätningen” kan samtidigt variera både för skrivbordskontroller och revisioner. Faktorer som kan ha betydelse är t.ex. komplexiteten hos det eventuella felet och utredningskompetens.

Vid val av effektmått bör också kostnaden för informationsinhämtning beaktas och vägas mot kvaliteten på effektmåtten (mer om detta i avsnitt 11.2). Det finns insatser där det inte, till en rimlig kostnad, går att identifiera effektmått med annat än låg kvalitet.³⁵ En naturlig fråga är hur man ska hantera effektmätningen för dessa insatser. Vi återkommer till denna fråga i avsnitt 11.1, men vi kan redan nu ställa oss frågan om det är meningsfullt att försöka *mäta* effekterna för denna av typ av insatser.

8.2 Att isolera effekter

8.2.1 Introduktion: Isolerbarhetskriterierna och det ideala upplägget

Förhoppningen att isolera effekten kan grusas av förändring i en mängd andra faktorer (de s.k. externa faktorerna i verksamhetslogiken) som samtidigt påverkar utfallet för den grupp som är föremål för insatsen, exempelvis:

- Andra pågående insatser
- Externa faktorer, såsom konjunktur och utbudschock
- Sociologiskt betingade beteendeförändringar
- Förändring av företagsledningen, såsom ny ägare eller revisor
- Dynamiska effekter, såsom effekter av tidigare genomförda kontroll- eller informationsinsatser³⁶

De operationaliserade effektmåtten för en insats kan i olika stor utsträckning påverkas av externa påverkansfaktorer. Låt oss titta närmare på två exempel, vilka har helt olika förutsättningar.

Exempel 1. En villaägare som bor nära en vältrafikerad väg blir nyfiken på hur stor del av ljudvågorna från vägen som ”stängs ute” av husets fönster. Utrustad med en decibelmätare mäter han ljudstyrkan vid ett av fönstren under eftermiddagsrusningen. Därefter öppnar han fönstret och mäter igen. Differensen mellan mätningarna före och efter fönsteröppningen är ett bra mått på hur mycket av ljudvågorna som fönstret ”stänger ute”.

Åtminstone fyra faktorer gör att villaägaren kan vara säker på att differensen är ett bra mått på fönstrets isolerande effekt:

³⁵ P.g.a. att det inte går, eller är mycket kostsamt, att finna operationella definitioner som rymmer mer än endast en eller ett fåtal aspekter av en innehållsrik teoretisk definition; eller att själva mätningen av effektmåtten inte kan göras med kvalitet.

³⁶ 2007 års temaavsnitt till Nordisk benchmarking, *Effektmål og effektmålinger*.

- Tiden för själva insatsen är kort

Villaägaren öppnar fönstret i det närmaste på nolltid.

- Effekten inträffar direkt efter insatsen

Ljudnivån förändras som en direkt effekt av att fönstret öppnas.

- Tiden mellan mätpunkterna är kort

Mätning av tillståndet (avläsning av decibelmätaren) före och efter insatsen sker i nära anslutning till insatsen.

- Insatsen är omfattande i förhållande till förändringarna i omgivningen

Den absolut primära ljudkällan är trafiken på vägen. Den är vid tiden för mätningen tung och fluktuerar endast marginellt. Andra påverkansfaktorer är i jämförelse små och/eller förändras lite över tiden. Fönsteröppningen förändrar dock påtagligt ljudnivån innanför fönstret.

Dessa fyra faktorer tror vi är centrala när det gäller att bedöma i vilken mån en ev. förändring av effektmåten är ett resultat av själva insatsen, d.v.s. i vilken mån det går väl att isolera effekten av själva insatsen genom en mätning före och efter insatsen, *utan* att använda någon jämförelsegrupp. Vi kallar de därför för *isolerbarhetskriterierna*.

Låt oss titta på ett mer komplext exempel.

Exempel 2. Skatteverket bedriver flera stora utvecklingsinsatser där flera år fortlöper från initiering till full implementering. Låt oss säga att olika delar av insatsen i fråga löpande implementeras, och att de två centrala målen med insatsen är att fullgörandekostnaderna ska minska och att förtroendet för Skatteverket ska öka.

De fullgörandekostnader man har som skattebetalare och det förtroende man känner för Skatteverket kan bero på en mängd olika faktorer (detta gäller sannolikt särskilt förtroendet, där förtroendet för institutioner och myndigheter i stort, skattesystemets utformning, etc. är exempel sådana faktorer). Förändringar av dessa faktorer kommer att ha en inverkan på förtroende och fullgörandekostnader. Komplexitetsgraden är därför stor. Tiden för insatsen är också lång. Effekterna uppstår också successivt, efter hand Skatteverket och skattebetalarna förstått och lärt sig tillämpa de nya systemen, reglerna, etc. Mätningarna är relativt dyra att göra, speciellt de som mäter fullgörandekostnader, och upprepas rimligen inte före och efter varje ny del av insatsen implementeras.

Sammantaget visar detta exempel att det är ytterst problematiskt att isolera effekterna av insatsen med avseende på förtroendet och fullgörandekostnaderna.

Ett par sätt att hantera dessa svårigheter står till buds. De operationaliserade effektmåten kan preciseras. Om det är en avgränsad del av fullgörandekostnaderna man särskilt försöker minska (och vilken påverkas i mindre omfattning av andra påverkansfaktorer) kan man försöka mäta förändringar i just denna del. Kanske är det också endast ett av implementeringsstegen som påverkar denna del av fullgörandekostnaderna – då kan mätningen göras före och efter detta steg.

Sammantaget visar de fyra *isolerbarhetskriterierna* när eventuella förändringar av effektmåten kan hänföras till insatsen, och när de inte kan det utan vidare. Kriterierna är särskilt lämpade att avgöra huruvida de *korta* effekterna enkelt kan mätas, eller ej. Om kriterierna inte

är uppfyllda måste vi alltså, om vi vill kunna mäta effekterna med god kvalitet, fundera över hur insatsens effekter kan isoleras från andra påverkansfaktorer. I vissa fall kan slutsatsen bli att det är nödvändigt att lägga upp insatsen på ett särskilt sätt för att detta ska vara möjligt. I andra fall kan slutsatsen bli att inte ens detta är möjligt, att det helt enkelt inte finns något bra sätt att isolera effekterna av själva insatsen.

Förutsättningarna att isolera effekterna av en insats är också ett centralt tema för den resterande delen av rapporten.

Det ideala upplägget för att isolera effekter

Det mest tillförlitliga sättet att studera effekterna av en insats är att välja ut två grupper som liknar varandra så mycket som möjligt i alla relevanta avseenden, och att därefter endast låta den ena gruppen, experimentgruppen, bli föremål för insatsen medan den andra, referensgruppen, inte blir det. Idealt sett motsvaras skillnaden i utfallet mellan grupperna av effekten av insatsen.

Med detta upplägg vill vi beskriva hur tillståndet skulle ha sett ut för den grupp som blev föremål för insatsen, om den inte blivit föremål för insatsen. Ett centralt antagande är alltså att referensgruppen och experimentgruppen påverkas i lika stor utsträckning, och på samma sätt, av förändringar i de externa påverkansfaktorerna. Här används alltså en referensgrupp för att beskriva den kontrafaktiska situationen.

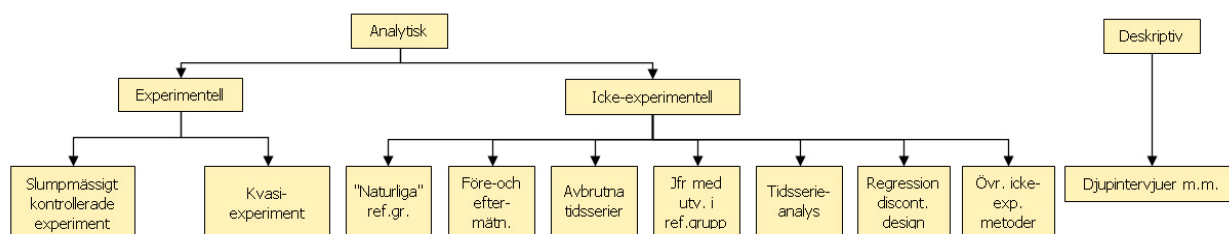
Att jämföra den grupp som blev föremål för insatsen med en referensgrupp är en mycket väl-etablerad metod för effektmätning. Den är dock varken möjlig eller önskvärd att använda för alla slags insatser, t.ex. kan insatsens målgrupp vara olik andra möjliga referensgrupper.

8.2.2 Klassificering av upplägg för att isolera effekterna

I fråga om möjligheter att isolera effekter av en insats vill vi ansluta oss till den s.k. evidenshierarkin som introducerades i kapitel 3.4.³⁷

En grov uppdelning utifrån evidenshierarkin säger att de experimentella uppläggen ger de bästa möjligheterna att isolera effekterna av insatsen, följt av de kvasi-experimentella och sist de icke-experimentella uppläggen. I figur 8.1 klassificeras de upplägg som presenteras och diskuteras i de kommande avsnitten av 8.2 (de deskriptiva metoderna diskuteras i kapitel 10).

Figur 8.1 Klassificering av uppläggen som behandlas i denna rapport



Observera att möjligheten att använda ovan nämnda upplägg beror på vilken slags insats som ska effektmätas samt för vilka information (operationaliserade effektmått) finns tillgänglig. Mer om detta i kapitel 9 – men låt oss först titta närmare på dessa upplägg.

³⁷ Därmed inte sagt att ambitionen med denna rapport är att plädera för ett upplägg framför ett annat. Vi vill visa vilka för- och nackdelar de innebär i termer av isolerbarhet, kostnader, etc.

8.2.3 Slumpmässigt kontrollerade experiment

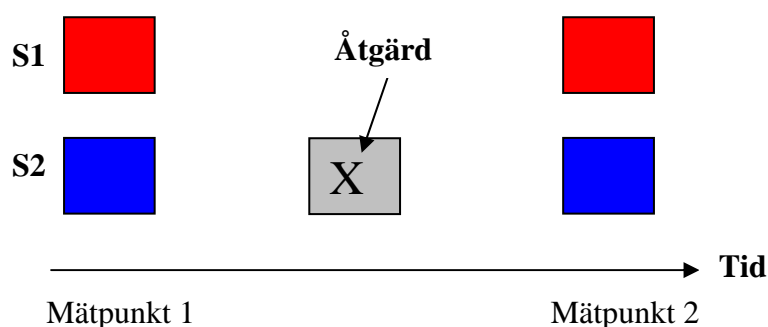
För att erhålla två likartade grupper är slumpmässiga urval det mest tillförlitliga.

Slumpmässiga urval. Om man vill undersöka en population (t.ex. hur till de röstberättigade skulle rösta om det vore riksdagsval idag, eller hur stor andel av skattebetalarna som gör ett felaktigt avdrag för övriga kostnader) genomförs undersökningen ofta endast för ett urval av populationen. Om populationen är stor blir undersökningen annars alltför kostsam. Om urvalet görs med hjälp av ”slumpen” blir stickprovet (de som väljs ut) representativt och undersökningen blir giltig för hela populationen. Ju fler som ingår i stickprovet, desto bättre beskriver det populationen och desto mindre blir osäkerheten i undersökningen.

Den vanligaste formen av slumpmässiga urval är *obundna slumpmässiga urval* (OSU) där alla element i populationen har samma sannolikhet att ingå i urvalet. I *stratifierade slumpmässiga urval* har *inte* alla element samma sannolikhet att ingå i urvalet, utan sannolikheten beror på vilken delmängd (stratum) av populationen som elementet ingår i.

För att mäta effekten av en åtgärd, X, är uppställningen i figur 8.2 ideal.

Figur 8.2 Idealt upplägg för effektmätning av en specifik åtgärd



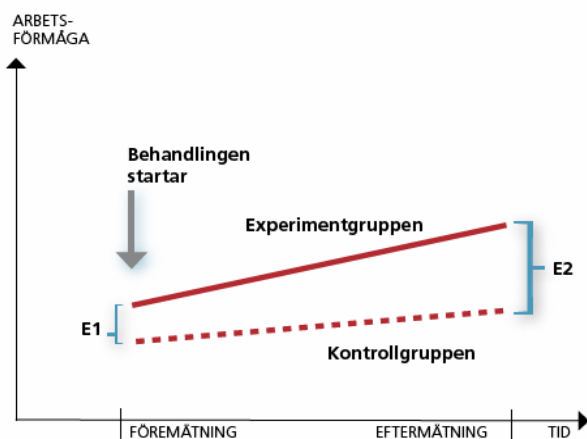
Ur den aktuella populationen genomförs en mätning inom två slumpmässigt utvalda grupper, S1 - referensgruppen - och S2 - experimentgruppen. Mätningen genomförs både *före* och *efter* insatsen. Genom upplägget isoleras effekten från andra påverkansfaktorer. Differensen i utfall motsvarar effekten av insatsen, med viss statistisk beräkningsbar osäkerhet. Effekten kan t.ex. uttryckas i termer av ett konfidensintervall³⁸.

I figur 8.3 ser vi att effekten av ett slumpmässigt kontrollerat experiment kan uppskattas genom att jämföra utvecklingen mellan experiment- och referensgruppen.³⁹ Estimerad (uppskattad) effekt av behandlingen på arbetsförmågan är en differens mellan två differenser, $E2 - E1$. Beräkningen kallas därför difference-in-difference.

³⁸ Ett 95 procentigt konfidensintervall anger att den ”sanna” effekten med 95 procent säsäkerhet ligger inom ett intervall kring det estimerade värdet, t.ex. mellan 4 och 12 procentenheter.

³⁹ Figur från ESV 2006:8. Observera att vi av uppenbara skäl valt att tala om referensgrupp, istället för kontrollgrupp.

Figur 8.3 Uppskattad effekt vid ett slumpmässigt kontrollerat experiment



Om man gör endast en *eftermätning* så sjunker säkerheten i effektmätningen något, men eftersom grupperna genom den slumpmässiga fördelningen i allt väsentligt är lika, kan effekterna estimeras och konfidensintervall beräknas. Detta upplägg är också att rekommendera om mätningen, genom t.ex. enkäter, med nödvändighet gör experiment- och referensgruppen medvetna om att de är föremål för extra uppmärksamhet. I dessa fall kan själva föremätningen resultera i ett förändrat beteende för såväl experiment- som referensgruppen.

Exempel. Våren 2007 utformades en pilotundersökning som ett slumpmässigt kontrollerat experiment i syfte att undersöka om betalningsbeteendet påverkades av tidigarelagda betalningspåminnelser om obetalade skatter. Två regioner ingick i studien. I Mälardalsregionen fördelades samtliga skattebetalare vid den aktuella avstämningstidpunkten slumpmässigt till att få antingen en betalningspåminnelse via brev eller telefon; i Stockholmsregionen fick ett slumpmässigt urval en betalningspåminnelse via brev.

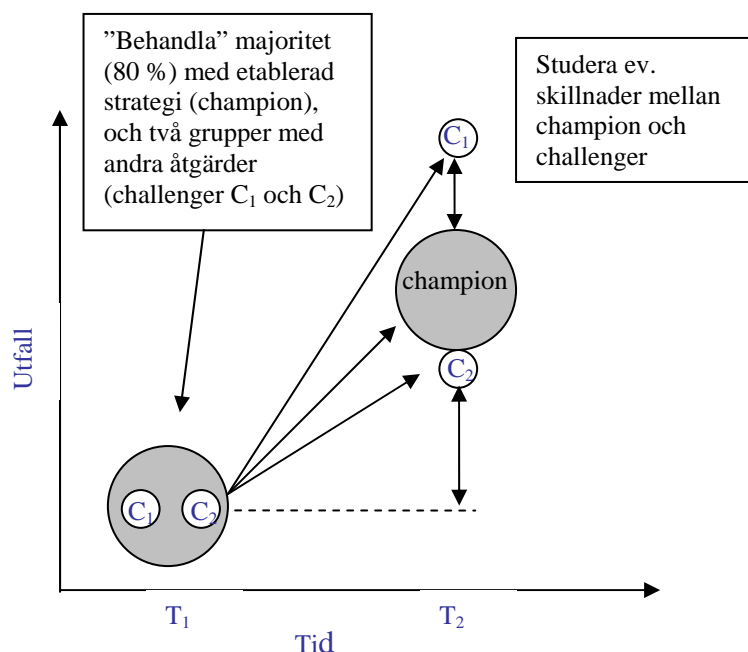
Genom att jämföra tid till betalning motsvarande den aktuella skatteskulden undersöktes hur betalningsbeteendet skilde sig åt mellan de tre aktuella grupperna (brev, telefonsamtal, ingen tidigarelagd påminnelse). Frågan var helt enkelt om tidigarelagd påminnelse medför att de obetalda skatterna betalas in tidigare än annars, och om den mer resurskrävande telefonkontakten är kostnadseffektiv i jämförelse med brevkontakt.

Resultatet av studien visade att effekten av en tidigarelagd betalningspåminnelse är god, både för telefon- och brevkontakt. I den grupp av skattebetalare som erhöll påminnelsen, hade en avsevärt större andel betalat sin skatteskuld efter 14 respektive 28 dagar, jämfört med gruppen som inte fick någon påminnelse. Något oväntat kanske gav studien inga belegg för att påminnelse via telefon gav större effekter på betalningsbeteendet än brevpåminnelsen. Effekten av brevpåminnelsen var istället, under studiens förutsättningar, statistiskt signifikant större än effekten av telefonpåminnelsen. (Voxberg, Danielsson & Blomkvist, 2007)

Fördjupning: Effektiva experimentupplägg

Studien i exemplet ovan illustrerar en tilltalade idé om hur de slumpmässiga kontrollerade experiment kan användas. Idén går under namnet *Champion/challenger method* och går i korthet ut på att nya strategier och arbetsmetoder, med liten risk, kan testas parallellt med existerande strategier (se figur 8.4).⁴⁰

Figur 8.4 Illustration av *Champion/challenger method*



Den stora majoriteten skattebetalare (eller motsvarande) blir föremål för den nuvarande strategin, medan ett mindre slumpmässigt urval blir föremål för en alternativ strategi, vilken bedöms ha potential att fungera bättre än den existerande. Så länge den alternativa strategin innebär ett potentiellt mervärde för skattebetalarna, kan man argumentera för att idén möjliggör ett kontinuerligt lärande utan etiska frågetecken. Om försöket resulterar i en ny *champion* kan motsvarande försök genomföras med nya utmanare, då dessa uppkommit.

Som experiment betraktat erhålls allra största kunskap om (1) elementen som ska ingå i undersökningen väljs ut slumpmässigt från *hela* populationen, och att det (2) bland dessa sker en slumpmässig fördelning till experiment- respektive referensgrupp. Under denna ideala situation blir resultatet av experimentet tillförlitligt och giltigt för *hela* populationen.

I samhällsvetenskapliga experimentella studier sker sällan (1), men definitionsmässigt (2). För en myndighet som vill arbeta med urval för åtgärd utifrån risk, kan det också tyckas ineffektivt och rättsosäkert att genomföra (1). Viktigt är dock att påminna sig om att resultatet av ett sådant experiment blir giltigt endast för den grupp som omfattas av studien.⁴¹

För att nå hög effektivitet i insatsen kan man utifrån vissa villkor, till exempel sådana som indikerar högre risk för fel, bestämma mot vilka insatsen ska riktas och därefter genomföra en slumpmässig fördelning.

⁴⁰ Figur med förlaga i Australian Taxation Office, 2007, s. 52.

⁴¹ För att göra generalisering till populationen plausibel görs undersökningen ofta om under olika förutsättningar och i olika delpopulationer. Ett bra exempel på detta är medicinska prövningar av nya behandlingsmetoder. Deltagarna i studien kan förstås inte slumpas fram, utan erbjuds att delta, väl medvetna om att de kan bli föremål för s.k. placebobehandling. Studier genomförs oftast på flera kontinenter etc.

Detta alternativ ger värdefull och säker kunskap och innebär, i normalfallet, endast något förhöjda kostnader (något lägre effektivitet och något högre administrativa kostnader).

Exempel på effektiva experimentupplägg. Låt oss säga att vi har identifierat en grupp företag med stor förväntad risk för fel. I bilaga 1 presenteras och diskuteras följande:

- Slumpmässig fördelning av bedömda högriskföretag till insats respektive inte insats
- Fördelning genom stratifierade slumpmässiga urval från hela populationen (där högriskföretag har stor sannolikhet för att bli föremål för insats, och lågriskföretag liten)
- Fördela tidpunkten för insats av högriskföretagen slumpmässigt i tiden.

I bilagan presenteras även ett experimentupplägg som är mer robust än det klassiska mot ev. avvikelser vid implementeringen. Detta rekommenderas vid större insatser som planeras med ett experimentellt upplägg.

Problem med slumpmässigt kontrollerade experiment

Som potentiella problem med slumpmässigt kontrollerade experiment har vi redan nämnt *mäteffekter* (själva mätningen påverkar beteendet). Ytterligare potentiella problem värda att nämna är *spridningseffekter* och *etiska problem*.

Medlemmar från referensgruppen kan direkt eller indirekt påverkas av insatsen riktad mot experimentgruppen. Under 2007 genomfördes en företagsanpassad informationsinsats inför KU-avstämningen. Insatsen var av pilotkaraktär och riktades till en avgränsad grupp företagare. För att undersöka effekterna av informationen undantogs, bland den avgränsade gruppen, en mindre slumpmässigt utvald grupp från informationen. Effektmätningen visade att effekterna var relativt små, och *spridningseffekter* lyftes upp som en tänkbar orsak till detta.⁴² Även om referensgruppen inte tilldelades informationen, kan informationskampanjen ha uppmärksamats och påverkat gruppens beteende.⁴³

Lämpligheten i att exkludera en grupp skattebetalare från en insats som är potentiellt nyttig för dem, har också diskuterats. Kunskapsbehovet avseende insatsens effekter måste här vägas mot principen om likabehandlingen (åtminstone i ett kortare perspektiv).

8.2.4 Kvasi-experimentell ansats

Det är inte alltid möjligt eller önskvärt att genom ett slumpmässigt urval fördela ”elementen” (t.ex. företag eller privatpersoner) till antingen experiment- eller referensgruppen. Ett alternativ kan då vara att genom ett strategiskt urval fördela dem på ett sådant sätt att de båda grupperna liknar varandra i de aspekter som är centrala för vilken effekt man kan förvänta sig av insatsen samt för den påverkan de externa faktorerna förväntas ha. Man talar då om en *matchad referensgrupp*. Denna ansats kan vara särskilt intressant att använda då populationen och/eller stickprovet är litet.

⁴² Danielsson & Voxberg (2007)

⁴³ En förvisso önskad effekt, som dock försvårar mätningen. Den eventuella spridningseffekten kan yttra sig genom att andra företagare och redovisningsrådgivare blir varse om att Skatteverket har en informationsinsats på området, vilket kan tolkas som att extra fokus på KU-avstämningarna och en fördjupad kontroll är att vänta. Spridningseffekten skulle då ha som effekt att även arbetsgivare som *inte* erhållit informationen, i större utsträckning redovisar kontrolluppgifter i överensstämmelse med skattedeklarationerna.

Matchningen bör ske utifrån i förväg definierade kriterier. Val av kriterier för urval bör bygga på välgrundad teori, vilken pekar ut de faktorer som förklarar variationen i den företeelse som studeras.⁴⁴

Exempel kan vara pilotprojekt som genomförs i några regioner eller på några kontor som helhet. Urvalet av de regioner eller kontor som genomför insatsen görs alltså aktivt, med syfte att kunna matcha dessa med liknande – i de för insatsen relevanta avseenden – regioner eller kontor, i vilka insatsen inte genomförs.

En kritik mot denna typ av studie är att man riskerar visa vad man antagit. Man kommer aldrig att med säkerhet kunna säga att en ev. skillnad mellan de båda grupperna är ett resultat av insatsen, en dålig teori och matchning, eller ”slumpen”. Vid ett slumpmässigt kontrollerat experiment erhålls ett mått på osäkerheten av effekten, något motsvarande erhålls inte vid en kvasi-experimentell studie.⁴⁵

8.2.5 Icke-experimentella metoder

Det är inte alltid möjligt eller önskvärt att *på förhand*, genom ett experimentellt eller kvasi-experimentellt upplägg, tillskapa en relevant jämförelsegrupp. Det kan dock finnas möjligheter att *i efterhand* identifiera referensgrupper, eller på annat sätt tämligen väl isolera effekterna av en insats. Om det finns ”naturliga” möjligheter att beskriva det kontrafaktiska tillståndet, med tillräcklig kvalitet, är vår utgångspunkt att dessa bör användas framför experimentella upplägg.

Detta område är mycket omfattande och nedan presenteras ett urval av dessa möjligheter. I bilagan introduceras också ett antal tämligen avancerade metoder.

8.2.5.1 ”Naturliga” referensgrupper

Vi inleder med en ganska generell kategori som vi kallar ”naturliga” referensgrupper. Vår förhoppning är de tankegångar som genomsyrar beskrivningen här ska uppmuntra ett kreativt synsätt och underlätta förståelsen av de övriga icke-experimentella metoderna.

I pilotstudien om de tidigare lagda betalningspåminnelserna användes, förutom de slumpmässigt utvalda referensgrupperna, även ”naturliga” referensgrupper i både tid och rum. Till exempel frågade vi oss hur betalningsbeteendet såg ut för de skattebetalare som hade skattekulder vid motsvarande avstämningstidpunkt förra året eller förra månaden. Likaså hur betalningsbeteendet såg ut innevarande månad för övriga regioner. Betalningsbeteendet i dessa referensgrupper liknade varandra och betalningsbeteendet i den slumpmässigt utvalda referensgruppen.⁴⁶ Detta är goda belägg för att dessa referensgrupper även ensamma skulle ha hållit mycket god kvalitet.

Om information är tillgänglig för samtliga skattebetalare i den aktuella populationen, kan det med en kreativ och analytisk ansats vara möjligt att finna goda ”naturliga” referensgrupper. Låt oss igen påminnas om syftet med referensgruppen, vilket är att beskriva den kontrafaktiska situationen.

⁴⁴ Vedung 1998, s. 148. Olika matchningsmetoder diskuteras i mer detalj i bilagan.

⁴⁵ Ett sätt att hantera detta kan vara att dela in populationen i grupper (s.k. strata) utifrån den typ av faktorer vi nämnde ovan. Därefter kan man med stratifierade slumpmässiga urval fördela objekten till antingen experiment- eller referensgrupp.

⁴⁶ Voxberg, Danielsson & Blomkvist (2007)

Principiellt finns det två dimensioner att jämföra i – i rummet och i tiden – vilket ovanstående exempel har illustrerat.

Vi ser att ”naturliga” referensgrupper kan identifieras som:

1. Liknande grupp, samma tid

Ex. Betalningsbeteendet i de övriga regionerna samma månad.

2. Samma grupp, annan/”liknande” tid

Ex. Restaurangernas redovisade löner föregående månad eller föregående år (se exempel personalliggare senare i 8.2.5).

3. Liknande grupp, ”liknande” tid

Ex. Betalningsbeteendet bland de som fick skatteskulder i samma region föregående månad eller samma månad föregående år.

Inom litteraturen brukar särskilt omnämnas *generisk kontroll*, vilken åsyftar en jämförelse med vad som är normalt hos en större, mer allmängiltig, grupp som inte blivit föremål för insatsen.⁴⁷

Ofta genomförs t.ex. urval för fördjupad kontroll utifrån ett riskperspektiv. Detta innebär att gruppen som blir föremål för insatsen inte är jämförbar med (representativ för) populationen i stort. Om man vill använda sig av någon slags generisk kontroll måste man göra en kvalitativ bedömning om i vilken mån grupperna är jämförbara.⁴⁸ Det faktum att ett företag redovisar t.ex. ovanligt låga intäkter kan bidra till att företaget blir föremål för fördjupad kontroll *och* att företaget har en stor potentialen att erhålla större intäkter nästkommande år. Att jämföra den kontrollerade och okontrollerade gruppen av företag med varandra, avseende utvecklingen av redovisade intäkter, kan därför vara mycket missvisande.

En metod som går under benämningen Regression discontinuity design erbjuder dock intressanta möjligheter att uppskatta effekterna av den utökade kontrollen, samtidigt som det finns ett antal tämligen avancerade matchningsmetoder som skulle kunna gå att använda för att estimerar effekten av revision (mer om dessa metoder senare under 8.2.5).

8.2.5.2 Före- och eftermätning

I avsnitt 8.2.1 presenterade vi de fyra isolerbarhetskriterierna som avgör i vilken mån det går att isolera effekten av själva insatsen vid en mätning före och efter insatsen, utan att använda någon jämförelsegrupp:

- Tiden för själva insatsen är kort
- Effekten inträffar direkt efter insatsen
- Tiden mellan mätpunkterna är kort
- Insatsen är omfattande i förhållande till förändringarna i omgivningen

Vårt exempel med villaägaren och fönstrens förmåga att isolera ljudet från den vältrafikerade vägen var förstas medvetet valt som ett idealtypiskt exempel. Men det finns verkliga exempel som i det närmaste uppfyller dessa villkor. En tämligen dramatisk förändring från en dag till en annan, t.ex. då en ny genomgripande lagstiftning träder i kraft, kan medföra att kriterierna ovan är väl uppfyllda.

⁴⁷ Se t.ex. Vedung 1998

⁴⁸ Värt att poängtera kan vara att vi nu inte kan använda statistiska metoder för att bedöma osäkerheten i effekterna (eftersom urvalen inte är slumpmässiga) utan den måste bedömas kvalitativt.

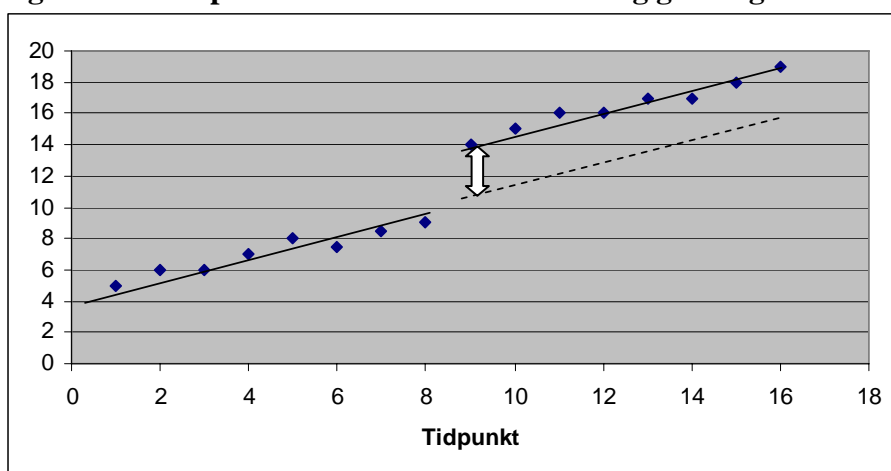
Om isolerbarhetskriterierna är väl eller ganska väl uppfyllda kan det vara tillräckligt att göra en före- och eftermätning.

Om vi kan följa utvecklingen av effektmåttet över tiden, genom upprepade mätningar eller registerdata, kan nedanstående ansatser vara värdefulla att använda (se vidare kap. 9.2 rörande olika informationstillgång). Om den aktuella populationen förblir oförändrad används lämpligen paneldata, d.v.s. mätning sker för *samma* objekt vid upprepade tillfällen. Om populationen förändras över tiden genomförs upprepade mätningar för ett tvärsnitt ur den för tillfället aktuella populationen (se exempel nedan).

8.2.5.3 Avbrutna tidsserier (eng: Interrupted time-series design)

Denna ansats är att föredra framför den enkla före- och eftermätningen. Om isolerbarhetskriterierna är väl uppfyllda kan förändringen av effektmåttet i stor utsträckning hänföras till själva insatsen. I figur 8.5 illustreras detta. Insatsen antas genomföras vid tidpunkt 8 och effekten inträffar direkt efter insatsen (effektens storlek illustreras av pilen i figuren).

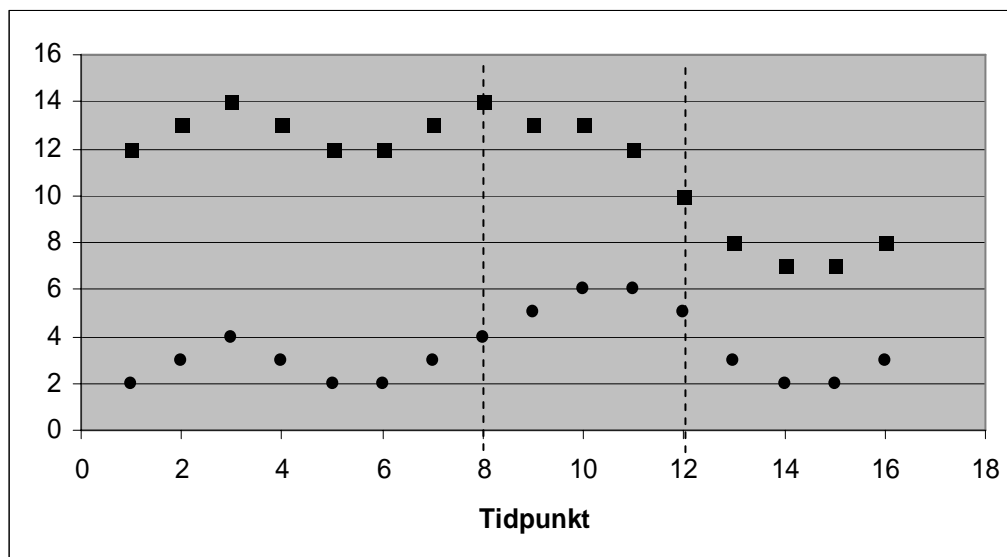
Figur 8.5 Exempel då en före- och eftermätning ger en god bild av effekterna av insatsen



8.2.5.4 Jämförelse med utvecklingen i en referensgrupp

I vissa fall kan det finnas en referensgrupp som förväntas utvecklas ungefär som gruppen som blivit föremål för insatsen. Notera att isolerbarhetskriterierna inte längre måste vara uppfyllda, det viktiga är att båda grupperna påverkas i ungefär samma utsträckning av de föränderliga externa faktorerna. Ju mer utvecklingen före och efter insatsen sammanfaller, desto bättre beskriver referensgruppen den kontrafaktiska situationen. I figur 8.6 illustreras detta. Förändringen av skillnaden (difference-in-difference) mellan grupperna används som mått på effekten. Insatsen påbörjas vid tidpunkt 8 och har nått full effekt vid tidpunkt 12. Effekten estimeras till 5 enheter, då differensen mellan grupperna minskat från 10 till 5 enheter.

Figur 8.6 En referensgrupp beskriver den kontrafaktiska situationen



8.2.5.5 Tidsserieanalys

I andra fall finns det externa faktorer som påverkar utvecklingen av effektmåttet, dock utan att det går att identifiera en referensgrupp som påverkas snarlikt av dessa externa faktorer. Ibland känner vi dessa faktorer och kan göra en prognos som tar hänsyn till dem. Med utgångspunkt i tidsseriens utveckling *före* tidpunkten då insatsen påbörjades kan s.k. tidsserieanalys användas för att beskriva den förväntade utvecklingen om insatsen *inte* skulle ha genomförts, d.v.s. det kontrafaktiska tillståndet.⁴⁹

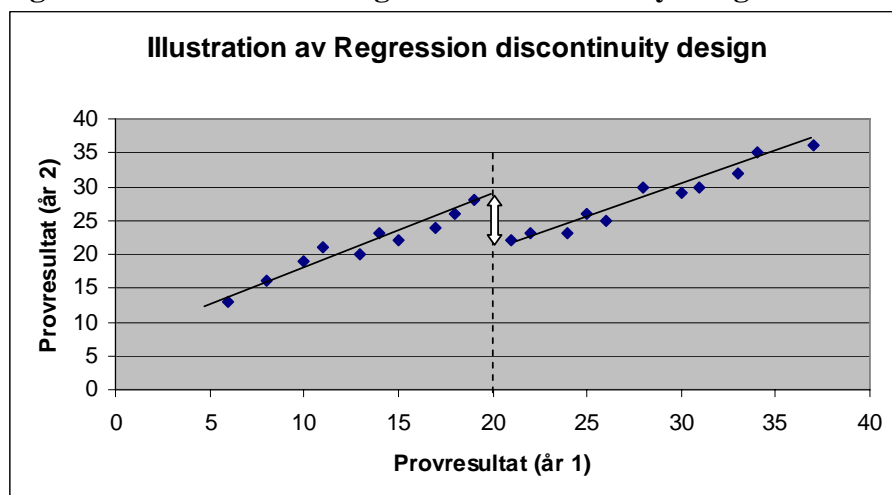
8.2.5.6 Regression discontinuity design

Detta är ett intressant upplägg som på ett enkelt, men elegant, sätt nyttjar den typ av villkor som ibland används för urval till insats. Upplägget kan användas då det finns en given brytpunkt på en kontinuerlig skala som avgör om t.ex. en person blir föremål för en insats eller inte. Brytpunkten fungerar ofta så att de för vilka behoven är som störst (t.ex. de elever med resultat på prov under en viss poäng) eller risken är högst (t.ex. de skattebetalare som gör kostnadsavdrag över en visst belopp), blir föremål för en insats – medan de andra inte blir det.

Låt oss säga att en grupp elever med lägre än 20 poäng på det senaste nationella provet i svenska får extraundervisning i ämnet under det efterföljande året. Resultatet på det nationella provet i samma ämne jämförs sedan med föregående års resultat. Man kan vänta sig att det finns en stark korrelation mellan elevernas provresultat de båda åren. Dessutom, om extraundervisningen har gett goda effekter kan man vänta sig att den grupp som hade strax *under* 20 poäng har utvecklats snabbare i ämnet än den grupp som hade strax *över* 20 poäng. Detta illustreras i följande spridningsdiagram (figur 8.7). Resultatet i exemplet är tydligt – för de eleverna som fick extraundervisning förbättras resultatet, i genomsnitt, mer än inom den grupp som inte fick extraundervisning.

⁴⁹ För en introduktion till tidsserieanalys, se exempelvis Newbold m.fl. 2003, kap. 17. För en fördjupad framställning kring analys av tidsserier och paneldata, se exempelvis Gujarati 2003.

Figur 8.7 Illustration av Regression discontinuity design



Det finns inga skäl att förvänta sig att den diskontinuitet som vi ser i spridningsdiagrammet i figur 8.7 skulle uppträda precis för den givna brytpunkten (provresultat år ett = 20), om inte extraundervisningen hade gett goda effekter.

Ur effektmätningssperspektiv är det idealt om brytpunkten är skarp, d.v.s. att samtliga på ena sidan av brytpunkten blir föremål för insatsen, och att ingen på den andra sidan blir det. Alternativet är t.ex. att det stora flertalet under brytpunkten blir föremål för insatsen, medan endast ett fåtal ovanför brytpunkten blir det.

Det finns goda skäl att förvänta sig att denna metod är lämplig att använda för att uppskatta de individualpreventiva effekterna av Skatteverkets utökade kontroller, inom ett antal områden.⁵⁰

8.2.5.7 Ett sammansatt exempel

Sammanfattningsvis har vi alltså studerat följande icke-experimentella metoder:

- ”Naturliga” referensgrupper
- Före- och eftermätning
- Avbrutna tidsserier
- Jämförelse med utvecklingen i referensgrupp
- Tidsserieanalys
- Regression discontinuity design

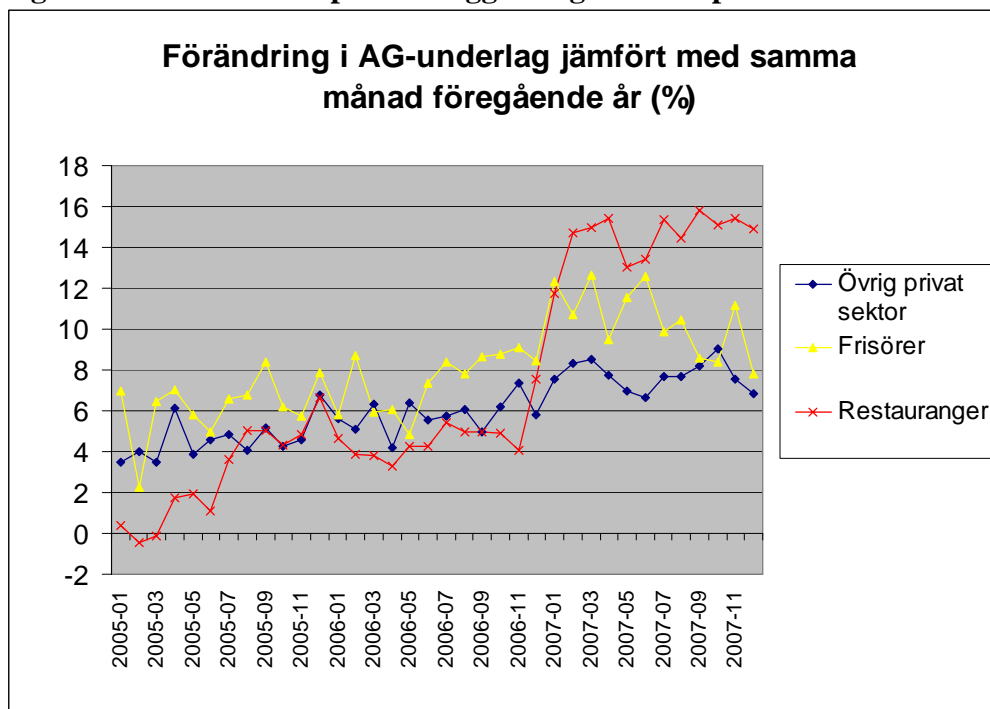
I exemplet nedan används i huvudsak ansats tre och fyra.

⁵⁰ Se vidare t.ex. HM Revenue and Customs 2008 för referenser till studier där denna metod har använts.

Exempel. Den 1 januari 2007 infördes lagstiftningen om personalliggare inom restaurang- och frisörbranschen. Det huvudsakliga syftet var att motverka s.k. svartarbete. Inom restaurangbranschen ökade också löneredovisningarna beloppsmässigt starkt kring årsskiftet. Højningen var bestående under året. Att hushållens ekonomi förbättrades under 2007 (minskad arbetslöshet, fler ungdomar har fått jobb m.m.) har troligen medfört ett i viss mån förändrat konsumtionsmönster till förmån för mer ”vardagslyx” i form av bland annat restaurangbesök. Den omedelbara ökningen av redovisade löner inom restaurangbranschen vid årsskiftet (se figur 8.8) talar dock enligt vår bedömning för att ökningen domineras av en effekt kopplad till personalliggarna.

Den dramatiska förändringen kring årsskiftet gör att de fyra isolerbarhetskriterierna i stor utsträckning bedöms vara uppfyllda. Flera referensgrupper har använts (i figuren nedan endast övrig privat sektor) och den totala effekten under året har estimerats som förändringen av differensen mot övrig privat sektor.⁵¹ Genom att jämföra redovisat AG-underlag med motsvarande månad föregående år, och genom att jämföra den branschspecifika utvecklingen med utvecklingen i den övriga privata sektorn, har vi således tagit hänsyn till den säsongsmässiga och – huvudsakligen – den konjunkturella variationen.

Figur 8.8 Införandet av personalliggarna gav effekt på redovisade löner



8.2.5.8 Fördjupning: Introduktion till ytterligare icke-experimentella metoder

Vi har sett att genom slumpmässiga urval kan experiment- och referensgrupp göras mycket lika – idealt sett är det endast förekomst av insats som skiljer grupperna åt. Det stora flertalet insatser genomförs dock med urval som inte är slumpmässiga (utan som en följd av självurval – t.ex. deltagande vid utbildningsinsats, eller urval initierat från myndigheten – t.ex. vid urval för informationsbesök eller revision). Vi har redan tidigare konstaterat att den resterande populationen i normalfallet inte utgör någon bra referensgrupp. Finns det då i dessa fall något sätt att ur den resterande populationen hitta en jämförbar referensgrupp?

⁵¹ Utöver *övrig privat sektor* har även andra, mer avgränsade och till branscherna mer liknande referensgrupper, använts i analysen, med liknande resultat. Figur från Danielsson & Franzon 2008.

I bilagan introduceras ett antal tämligen avancerade metoder som i vissa fall kan besvara denna fråga jakande. Följande metoder diskuteras:

- Kvalitativ eftermatchning
- Propensity score matching
- Multivariat matchning
- Genetisk matchning
- Instrumentvariabelmetoden

En gemensam förutsättning för matchningsmetoderna är att observerbar *och* tillgänglig information väl kan beskriva vilka som blev föremål för insatsen. Vidare måste sannolikheten att bli föremål för insatsen vara större än 0, men mindre än 1, för samtliga i den aktuella populationen. I allmänhet rekommenderas att dessa metoder – med undantag för kvalitativ efterhandsmatchning – används endast vid relativt omfattande och strategiskt viktiga insatser. Den individualpreventiva effekten av revision skulle exempelvis troligen kunna studeras med hjälp av någon av dessa matchningsmetoder.⁵²

Vi ska strax återkomma till de olika upplägg som introducerats i detta avsnitt och föra samman dem med de förutsättningar som råder för den typ av insatser som Skatteverket genomför eller har att kontrollera efterföljelsen av. Först ska vi dock se för vilka resultatet av en effektmätning är giltigt.

8.3 Generaliserbarhet: för vilka är resultatet giltigt?

I all enkelhet kan sägas att resultatet, om mätningen är riktigt genomförd, är giltigt åtminstone för de som omfattas av effektmätningen, men endast under vissa förutsättningar för någon större grupp.

För insatser där urval görs, gäller att endast ett slumpmässigt urval kan sägas representera något mer än sig självt (se 6.2.1). Om det t.ex. handlar om en riktad kontroll, är resultatet av t.ex. en mätning av den individualpreventiva effekten giltigt endast för denna grupp. Att generalisera resultatet till att gälla en större grupp är i allmänhet osäkert. Om vi istället genomför ett klassiskt slumpmässigt kontrollerat experiment (där objekten väljs ut slumpmässigt från hela den aktuella populationen) blir resultatet giltigt, med viss beräkningsbar osäkerhet, för hela den aktuella populationen.

Om det handlar om en insats som inte är baserad på ett urval, t.ex. en insats som kan tänkas ge allmänpreventiva effekter (som t.ex. införandet av ny lagstiftning), är resultatet av mätningen giltigt för hela den aktuella populationen.

När man inom kvalitativ metod talar om generaliserbarhet brukar man tala om studier som ”fall” av något större, mer generellt. Det kan t.ex. vara värdefullt att fråga sig om resultatet av en effektmätning skulle kunna säga något om i vilken mån motsvarande arbetsmetod, angreppssätt, etc. skulle kunna ge goda effekter mer generellt.

⁵² Urval för revision kan bedömas uppfylla förutsättningarna för matchningsmetoderna och en studie om effekterna av revision skulle dessutom kunna ge svar på frågorna om vad som kännetecknar de företag som blir föremål för revision, och vad som i sin tur kännetecknar de företag som enligt genomförd kontroll har undandragit skatt.

9. Att mäta effekter – utifrån en skatteförvaltnings förutsättningar

I detta kapitel ska vi titta närmare på de särskilda förutsättningar som gäller för effektmätning av de insatser som Skatteverket genomför. I stort är framställningen allmänt giltig, men Skatteverkets insatser har några särdrag som gör att vi kan precisera och sätta en del ramar kring den generella teori som diskuterades i föregående kapitel. Särskilt kommer möjliga upplägg att diskuteras utifrån kategori av insats (9.1) och informationstillgång (9.2).

9.1 Olika upplägg möjliga för olika kategorier av insatser

En vanlig kategori av insatser är de där ett *urval av skattebetalare* blir föremål för insatsen. Skatteverkets fördjupade kontroll (s.k. skrivbordskontroll eller revision) och riktade informationsinsatser är exempel på denna kategori av insatser. Vanligen blir ett stort antal skattebetalare föremål för denna insats, samtidigt som ett stort antal skattebetalare ur samma population inte blir föremål för den aktuella insatsen, den aktuella perioden. Det finns således *många* i insatsgruppen och *många* i en möjlig referensgrupp.

Det finns också insatser som är lika för hela den aktuella populationen, t.ex. en lagstiftningsförändring som gäller samtliga företagare. Någon möjlighet att välja ut vilka grupper som ska bli föremål för insatsen och inte, finns ej. Antalet i en eventuell referensgrupp är 0.

Ovanstående båda exempel motsvarar diametralt olika förutsättningar när det gäller *möjligheten* att välja upplägg för en effektmätning. Tabell 9.1 illustrerar sambandet mellan antalet i insats- respektive referensgruppen och möjliga upplägg (se avsnitt 8.2). Observera att enheterna som räknas och kategoriseras (1, få, många) kan vara skattebetalare, regioner, branscher, etc. Karaktären på insatsen sätter alltså sina begränsningar på hur effektmätningen kan genomföras.

Tabell 9.1 Sambandet mellan insatsens karaktär och möjliga upplägg

		Antal i insatsgr.		
		1	få	många
Antal i ref.gr.	0			
	få			
	många			

Möjliga upplägg enligt tabell 9.1 är:

1. Vit: Före- och eftermätning, avbrutna tidsserier, tidsserieanalys
2. Ljusgrå: Som ovan + ”naturliga” referensgrupper + jämförelse med utvecklingen i referensgrupp
3. Mörkgrå: Som ovan + kvasi-experiment
4. Svart: Som ovan + regression discontinuity design (under preciserade villkor) + slumpmässiga kontrollerade experiment + övriga icke-experimentella metoder

Detta innebär alltså att insatser som tillhör kategori 4 ger de största *möjligheterna* att *välja* upplägg som isolerar effekten av insatsen, medan insatser som tillhör kategori 1 inte ger några möjligheter för detsamma.

Ett par exempel:

- Om ett antal branscher, regioner (eller andra väldefinierade enheter) blir föremål för en insats, medan andra inte blir det, finns det *få* i insatsgruppen och *få* i en möjlig referensgrupp. Möjliga upplägg för dessa slags insatser framgår av

kategori 1, 2 och 3 ovan (Före- och eftermätning, avbrutna tidsserier, tidsserieanalys, ”naturliga” referensgrupper, jämförelse med utvecklingen i referensgrupp, samt kvasi-experiment)

- Insatser av typen utökad kontroll, riktade informationsinsatser, etc. faller inom ”många” – ”många”-cellen i tabellen ovan. Detta innebär att samtliga upplägg är möjliga att genomföra för denna typ av insatser.

9.2 Möjliga upplägg utifrån informationstillgång

I detta avsnitt kommer vi att studera vilka upplägg som kan vara aktuella för Skatteverkets insatser, beroende på för *vilka* information (operationaliserade effektmått) finns tillgänglig.

Förutom kategori av insats ser vi nämligen att informationstillgången i allra högsta grad påverkar vilka upplägg som är möjliga och, mer generellt, hur effektmätningen kan genomföras. Utifrån detta perspektiv ser vi att den data som Skatteverket har, eller kan göra, tillgänglig för en insats kan sägas tillhöra någon av följande tre kategorier:

1. Information av god kvalitet finns tillgänglig för hela populationen
2. Information av god kvalitet kräver mätning genom insats (kontroll)
3. Information kan tillgängliggöras för ett urval av populationen (dock utan att mätning sammanfaller med insats/kontroll)⁵³

Man kan tänka sig två ytterligheter i fråga om kostnad, kvalitet och omfattning av information:

- Information erhålls löpande (och gratis), av god kvalitet och för samtliga relevanta skattebetalare i populationen
- Information erhålls som resultat av insats, för ett urval av skattebetalare, och av låg kvalitet

9.2.1 Information av god kvalitet finns för hela populationen

Def: Information erhålls med god kvalitet, vilken väl återspeglar de verkliga effekterna, för samtliga i populationen.

Om information av god kvalitet finns tillgänglig för hela populationen innebär informationstillgången ingen begränsning. Sambandet mellan typ av insats och möjligt upplägg som illustreras i tabell 9.1 gäller. Dessutom är resurserna som krävs för själva effektmätningen relativt små.

Exempel kan vara information från deklarationsblanketterna (och annan för Skatteverket tillgänglig registerdata). Deklarerade löner i fallet personalliggarna är ett bra exempel som vi redan nämnt. I en studie över effekten av kontrollen av avdragsyrkanden för övriga kostnader under inkomst av tjänst, studerades förekomst och storlek av avdragsyrkandet nästföljande år.⁵⁴ Informationen är den skattebetalaren har lämnat, så i den meningen är kvaliteten perfekt. Frågan är således i vilken mån informationen återspeglar den effekt man vill studera.

⁵³ Observera att man kan tänka sig att använda kontroll för att mäta regelefterlevnad inom en population av skattebetalare, före och efter t.ex. en omfattande informationskampanj. Då används kontrollen som mätmetod, men sammanfaller ej med själva insatsen. Detta fall tillhör kategori tre.

⁵⁴ RSV rapport 2001

En del av skattebetalarens interagerande med Skatteverket låter sig också väl studeras. Det kan handla om företagsregistreringar och avslut, tidpunkt för inlämning av deklarerationer, betalningsbeteenden, m.m.

Insatser som avser att påverka faktorer av den här karaktären har alltså en god informationstillgång.

9.2.2 Information av god kvalitet kräver mätning genom insats (kontroll)

En stor del av Skatteverkets verksamhet består av kontroll som insats. I vissa lägen krävs också kontroll för att mäta graden av regelefterlevnad efter genomförd kontroll (som insats), andra möjliga effektindikatorer som finns tillgängliga genom t.ex. deklarationsuppgifter anses inte hålla tillräckligt god kvalitet.

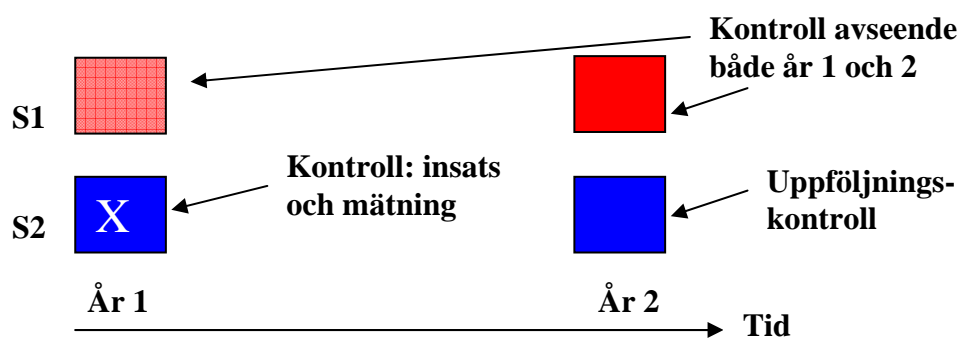
Def: Information av god kvalitet, efter genomförd kontroll (som insats), kräver också mätning genom kontroll.

Under dessa förutsättningar är inte det ideala upplägget som beskrevs i figur 8.1 tillämpbart. Problemet är att vi inte kan genomföra en mätning av de som blivit föremål för insatsen utan att dessa påverkas av mätningen.⁵⁵ *Mätningen före* och själva *insatsen* sammanfaller under dessa förutsättningar med nödvändighet. Det är därför inte möjligt att mäta effekten av kontroll med riktigt samma upplägg, om mätningen i sig själv kräver kontroll.

Det finns några tänkbara lösningar på denna problematik:

1. Vi börjar med det, ur isolerbarhetsperspektiv, ideala upplägget. För referensgruppen, S1, genomförs kontroll samtidigt avseende både uppgifterna år 1 och år 2. På detta sätt kan vi med säkerhet uttala oss om vad som hänt med gruppen S1 utan att denna påverkas av den insats som vi vill mäta effekten av.

Figur 9.1 Idealt upplägg om effektmätning av genomförd kontroll också kräver mätning genom kontroll



Den eventuella differensen (d_1) mellan år 2 och år 1 för grupp S1 ger ett mått på storleken av de externa påverkansfaktorerna (inkl. den allmänpreventiva effekten av kontroll). För grupp S2 kan differensen mellan år 2 och 1 betecknas d_2 . Skillnaden mellan den individualpreventiva effekten på S2 och den allmänpreventiva effekten på S1 kan då uttryckas som $d_1 - d_2$.

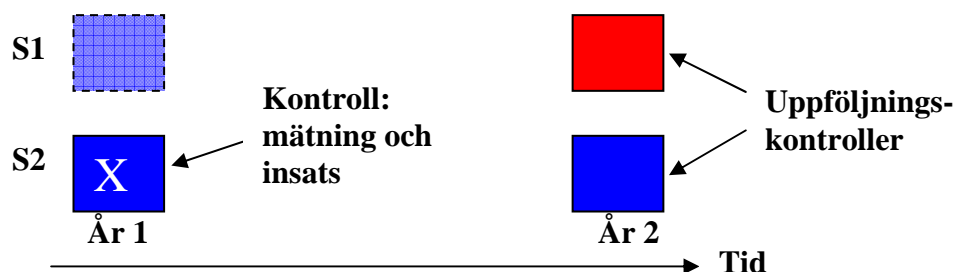
Även om ovanstående upplägg kan vara önskvärt kan det finnas legala, och etiska, hinder mot att genomföra det. Den period som Skatteverket får genomföra s.k. omprövningar inom

⁵⁵ Begränsningen är densamma som vid studier av subatomära partiklar.

sträcker sig nämligen ett år framåt. Därefter måste Skatteverket visa att skattebetalaren har lämnat oriktig uppgift, för att kunna fatta ett beslut som är till skattebetalarens nackdel. Det betyder att upplägget i figur 9.1 är tillåtet, men att det inte möjligt att utvidga det till att omfatta mer än två år.

2. Sedan ett alternativ som liknar det vi presenterade som endast eftermätning.

Figur 9.2 Upplägg som liknar endast eftermätning



Med hjälp av uppföljningskontroller kan vi med tämligen god säkerhet uttala oss om den individualpreventiva effekten av kontrollen. Om kontrollinsatsen är omfattande kan dock gruppen S1 påverkas så att differensen svarar mot den individualpreventiva effekten minus den allmänpreventiva effekten på S1.

3. Om kontrollen inriktas mot en förväntad högriskgrupp kan skattebetalarna fördelas till antingen experiment- eller referensgrupp. För övrigt kan alternativ 1 eller 2 ovan tillämpas. (Se bilaga för motivering och fler exempel på möjliga upplägg.)
4. Till sist, ett alternativ som är jämförbart med en före- och eftermätning. Om insatsen är omfattande i förhållande till förändringarna i omgivningen (övriga påverkansfaktorer är små och/eller förändras lite över tiden) kan helt enkelt uppföljningskontroller genomföras.

Då detta informationsalternativ gäller är effektmätningen som mest resurskrävande. Detta gäller särskilt om revision anses vara den nödvändiga kontrollformen (jfr avsnitt 6.2.2). Att genomföra uppföljningskontroller – med bl.a. ökade fullgörandekostnader som följd – hos t.ex. företag utan anmärkning vid första kontrolltillfället, med anledning av ett experimentellt upplägg, är också etiskt tveksamt.

Observera att dessa upplägg är aktuella endast om kontroll anses nödvändig för att mäta regel- efterlevnaden, efter tidigare genomförd kontroll, med god säkerhet. I de fall där effektmått av god kvalitet finns tillgängliga för hela populationen, kan vi arbeta med samtliga tillgängliga upplägg. Observera också de möjligheter som framför allt metoden Regression discontinuity design ger att mäta effekten av den utökade kontrollen.

9.2.3 Information kan tillgängliggöras för ett urval av populationen

Def: Information av god kvalitet erhålls ej för samtliga i populationen, men information som svarar väl mot effekten kan inhämtas utan att mätning och insats sammanfaller.

Exempel på denna metod för datainsamling kan vara enkätundersökningar, tredjemansuppgifter eller kontroller (då det ej handlar om en kontrollinsats som ska effektmätas).

Så länge informationsinhämtningen inte ”krockar” med själva insatsen är alla föreslagna upplägg, i princip, möjliga. Om isolerbarhetskriterierna är väl uppfyllda kan två mätningar – en före och en efter en insatsen – göras för att teckna en bild av insatsens effekter. En enkätundersökning före insatsen bör, i normalfallet, omfatta ett slumpmässigt urval ur den relevanta populationen. Om risk för mäteffekter föreligger, bör undersökningen efter insatsen omfatta ett nytt slumpmässigt urval. Om det är en serie av händelser, eller en utveckling över tid, som ska studeras kan s.k. panelurval användas. Mätning sker då, med en given periodicitet, genom obundna slumpmässiga urval, samtidigt som urvalsobjekten stegvis byts ut över tiden.

Resurser för informationsinhämtning – för Skatteverkets och för skattebetalarnas del – kan dock vara betydande. Risken för mäteffekter och låg validitet i mätningen föreligger också om t.ex. en enkätundersökning genomförs såväl före som efter en insats tillhörande kategori 4 (t.ex. insatser med urval för kontroll och riktade informationsinsatser). Mätningen riskerar nämligen att direkt påverka beteendet om skattebetalarna blir varse att de är föremål för extra uppmärksamhet. Om ett experimentellt upplägg används räcker det väl, som vi sett, att göra en eftermätning för både experiment- och referensgruppen.

10. Att belysa effekterna av en insats med kvalitativ metod

Här ska kort sägas något om vad som skiljer kvalitativa ansatser från de kvantitativa, och något om hur sådana kan utföras i vår verksamhet. Metodlitteraturen är oöverskådligt omfattande, och referenslistan ger några exempel för vidare läsning.

En kvalitativ ansats lämpar sig inte lika bra som en kvantitativ för att isolera variabler och kanske inte heller för att pröva redan ställda hypoteser. Dess styrka ligger istället i att upptäcka nya fenomen eller förklaringar, eller att skapa förståelse om de undersökta företeelserna eller personerna. Kvalitativa metoder kan hjälpa oss att formulera teorier om varför olika åtgärder har effekt på vår omgivning, teorier som sedan kanske kan prövas med kvantitativa metoder.

Kvalitativa metoder är användbara när det är värdefullt att låta människor komma till tals om våra insatser: skattebetalare, revisionsbyråer, branschorganisationer etc., eller när vi behöver analysera värderande eller beskrivande texter. Detta kan bero på att det inte går att göra en kvantitativ effektmätning på grund av databrist, -kvalitet eller urvalsproblem etc. Eller vill vi redan från början ha svar på frågor som inte går att besvara med kvantitativa metoder, t.ex. *varför* vår insats har haft effekt (eller inte). Men man kan med god säkerhet säga att en kombination av kvantitativa och kvalitativa metoder ofta ger det bästa resultatet. Vårt förslag på arbetssätt med en utbyggd verksamhetslogik kan t.ex. sägas vara en kvalitativ metod.

Det bör emellertid poängteras att kvalitativa metoder inte följer samma rigorösa mönster för genomförande som de kvantitativa, eftersom det inte är laboratoriesituationen som är förebilden utan snarare det goda, informativa samtalet. Resultaten är därför inte heller generaliserbara i samma utsträckning som vid kvantitativa mätningar.

En variant av blandad metod, som innefattar såväl kvantitativa som kvalitativa moment är *skuggkontroll*. Detta innebär att man låter andra experter, administratörer, branschorganisationer eller dylikt uttala sig om mätdata före respektive efter en insats, om deras bedömning av hur resultatet hade blivit om insatsen inte hade vidtagits. Metoden är en variant av före-/efter-mätning, där man blandar in andra i tolkningen av det kontrafaktiska tillståndet.

I de fall vi skulle vilja veta varför en insats har haft effekt, eller inte haft effekt, finns möjligheten att diskutera detta med dem som utsatts för insatsen. Man kan överväga att utföra en enkätstudie eller intervjuer. Märk att enkäter kan vara både kvantitativt och kvalitativt inriktade, beroende på hur dessa utformas. Om man utför ett slumpmässigt urval på ett korrekt sätt och har ett frågeformulär med fasta svarsalternativ, används enkäten som en kvantitativ metod. Om man istället t.ex. delar ut enkäten till alla som utsatts för en kontroll, och svarsalternativen i stor utsträckning är öppna (dvs. tomma fält för den svarande att själv fylla i) används enkäten istället som en kvalitativ metod. Användningsområdet för dessa skiljer sig också åt. I det förra fallet kan man förvänta sig att göra generella uttalanden om hela den grupp som de svarande representerar, vilket man inte kan med det senare fallet. Öppna svarsalternativ är att föredra t.ex. när vi har utfört något för första gången, och inte vet vilken reaktion vi kan förvänta oss.

Givetvis finns metodproblem även med dessa ansatser, t.ex. bör vi nog bara förvänta oss uppriktiga svar från dem som verkligen vill samarbeta i en sådan undersökning. Detta kan kanske vara svårt i vissa fall, t.ex. vid vissa kontrollprojekt etc. Dessutom är såväl enkätundersökningar som intervjustudier resurskrävande. Om enkäter eller intervjuer övervägs, bör man noga fundera på var kompetensen finns att utföra detta. I fråga om enkäter bör rimligen analysenheten på huvudkontorets produktionsavdelning kontaktas, som utför Skatteverkets större enkätundersökningar.

Det är inte alltid möjligt att få information direkt från dem som varit föremål för våra insatser, t.ex. av etiska skäl eller av kostnadsskäl. Då kan man överväga att fråga intresseorganisationer (om sådana finns) om deras och deras medlemmars uppfattning om Skatteverkets insats. Ett konkret exempel på detta gjordes i samband med effektmätningen av insatsen med personalliggare. Resultatet av den kvantitativa mätningen, som visade på effekt genom högre redovisade AG-underlag inom restaurangbranschen jämfört med föregående år, diskuterades tillsammans med deras branschorganisation, SHR, för att få deras syn på bl.a. Skatteverkets insats. Vid ett möte med SHR fördes en diskussion kring frågor som organisationen hade fått någon vecka före mötet.

Figur 10.1 Frågeutskick till branschorganisation

Erfarenheter från SHR i fråga om personalliggare

För utvärdering av vårt arbete är det av stor vikt med synpunkter från berörda branschorganisationer. Vi vill därför gärna ha svar på följande frågor.

1. Vilka är Era erfarenheter av fullgörandekostnader för företagen?
2. Hur har bemötandet från Skatteverket varit vid de kontroller som utförts?
3. Har ni uppfattat regionala skillnader i kontrollen?
4. Vad kunde Skatteverket gjort bättre?
5. Uppfattar Ni att införandet av personalliggare lett till en sundare konkurrens i branschen?
6. Andra synpunkter?

Om vi inte har möjlighet att få information från vare sig de som utsatts för insatsen eller intresseorganisationer, kan man överväga om det går att göra någon form av benchmarking, dvs. jämföra våra erfarenheter med andra (skatte-)förvaltningars.

11. Avslutande diskussion

I detta kapitel knyter vi ihop tankarna i del III genom att diskutera under vilka förutsättningar som det är meningsfullt att försöka mäta effekter av en insats, vad som kan vara vägledande för beslutet att lägga extra resurser på en effektmätning (när detta krävs), och hur resultatet från en effektmätning lämpligen tas om hand.

11.1 För vilka insatser är det meningsfullt att mäta effekter?

Vi har sett att för vissa insatser kan det vara i det närmaste omöjligt att finna lämpliga effektmått och/eller att isolera effekten av själva insatsen från effekten av andra påverkansfaktorer. Detta ställer oss inför den delikata frågan; för vilka insatser är det meningsfullt att försöka *mäta* dess effekterna?

När Ekonomistyrningsverket diskuterar val av indikatorer för måluppfyllelse framhåller man vikten av att inte vara ”perfektionistisk och intolerant”, mått som är bara är lite relevanta bör inte avfärdas förrän bättre mått har presentats.⁵⁶ Vi delar perspektivet att val av indikatorer kan vara en kontinuerlig utvecklingsprocess. Vi ställer oss dock tveksamma till användning av indikatorer som rymmer endast en eller ett fåtal av de mångfacetterade aspekter av de effekter som man vill uppnå med insatsen. Om man väljer att använda dessa effekttindikatorer finns det en risk att effektmätningen blir vilseledande (och att insatsen styrs fel, jfr avsnitt 6.2.6).

⁵⁶ ESV 2006:7, s. 17

Detsamma gäller om det inte är möjligt att ens ungefärligt isolera effekten av själva insatsen från andra påverkansfaktorer.

Vi menar att det är meningsfullt att försöka mäta effekter då kvaliteten på effektmåtten är god eller ganska god *och* det går väl eller ganska väl att isolera effekterna av insatsen, men inte annars.

Om detta villkor är uppfyllt eller inte, för en given insats, kan bero på de resurser som man är beredd att lägga på själva effektmätningen. Vi återkommer till denna fråga i nästa avsnitt.

Vad är då alternativet till att *mäta* effekter med kvantitativ metod, om vi konstaterat att det givet insatsens utformning och informationstillgång inte är meningsfullt att mäta effekterna? Vår rekommendation är då att effektutvärderingen genomförs med utgångspunkt i den utökade verksamhetslogiken (7.2) och med hjälp av kvalitativ metod (kap. 10) – för att belysa effekterna – eller inte alls.

11.2 För vilka insatser är det värdefullt att lägga extra resurser för att kunna mäta dess effekter?

Effektmätning innebär i sig själv nyttor och kostnader för organisationen. Det handlar såväl om kostnaden för själva mätningen som, i föreliggande fall, extra kostnader för informationsinhämtning och något slags experimentupplägg som kan medföra att insatsen blir mindre effektiv (se avsnitt 8.2.3 samt bilaga för diskussion kring upplägg som kan vara förenliga med en effektiv verksamhet).

Det kan förstås finnas krav från regeringens sida att myndigheten gör en bedömning av vilka effekter dess insatser ger. I detta avsnitt tar vi dock enbart sikte på de interna behoven av effektmätning som finns.

- Bland de insatser som man överväger att effektmäta är frågan om de uppfyller något av villkoren som gör att effekterna av insatsen kan mätas med god kvalitet, till en relativt låg kostnad, i ett icke-experimentellt upplägg (t.ex. att de uppfyller isolerbarhetskriterierna eller att det finns "naturliga" referensgrupper).
- Bland de resterande insatserna är frågan om det *är möjligt* att genomföra något experimentupplägg och/eller inhämta ytterligare information, vilket möjliggör en effektmätning av god kvalitet.
- Frågan till sist är för vilka insatser det är motiverat att lägga extra resurser på effektmätningen.⁵⁷

En effektmätning är lite värd, oavsett kvalitet, om den inte fyller ett reellt kunskapsbehov rörande insatsens effekter. Vi behöver bedöma nyttan och kostnaden med att genomföra en effektmätning med större krav på *upplägg* och/eller *informationsinhämtning*.

För de insatser där det skulle krävas betydande resurser att möjliggöra en effektmätning av god kvalitet, är alltså frågan i vilka fall detta är värdefullt att göra? Vi ser att dessa ställningstaganden bör vägledas av tre frågor:

- 1 Vilken nytta skulle det innebära att mäta effekterna av insatsen med god kvalitet?

⁵⁷ Ta gärna hjälp av *Flödesschema för effektmätning* för att följa vägen fram till denna beslutspunkt.

Nyttan av att mäta effekterna av olika insatser kan bero på en mängd faktorer. För en lärande organisation handlar det om behovet av kunskaper om i vilken mån insatserna leder till goda effekter. För vissa insatser är denna kunskap redan stor, för andra är denna kunskap liten – men av yttersta vikt, då strategiskt viktiga beslut fattas med utgångspunkt i hur väl dessa insatser ger de önskade effekterna.

Skatteverket genomför en rad insatser inom ramen för sin riskhantering. För denna typ av insatser vill vi föreslå ett antal kriterier för att värdera *nyttan* med genomförd effektmätning:

- Risker och möjliga åtgärder inom det riskområde som insatsen sker
 - Riskerna inom området är allvarliga
 - Hittills vunna erfarenheter tyder på att riskerna är svårhanterliga
 - Det finns många åtgärder mellan vilka man måste prioritera
 - Den empiriska kunskapen om hur väl dessa olika åtgärder fungerar (vilka effekter de ger) är låg
- Insatsens syfte och resursåtgång
 - Syftet med insatsen är helt eller delvis att utveckla kunskapen om effektiva åtgärder/arbetssätt
 - Insatser föranledda av ny lagstiftning eller dyl.
 - De resurser som insatsen tar i anspråk är stora
- Bidrag till ”helheten”
 - Resultaten från effektutvärderingen kan säga något om i vilken mån motsvarande arbetsmetod, angreppssätt, etc. skulle vara effektiva mer generellt (och således bidra till att utveckla Skatteverkets strategier).

2 Vilket mervärde innebär det att, om möjligt, välja ett *upplägg* som bättre kan isolera effekterna av insatsen? Vilka kostnader är de olika uppläggen associerade med?

Grundläggande handlar det om i vilken mån den aktuella insatsen och informationstillgången ger möjlighet att välja upplägg (se avsnitt 9.1 och 9.2) och i vilken mån de fyra isolerbarhets-kriterierna (se 8.2.1) är uppfyllda, eller det är möjligt att på icke-experimentell väg beskriva det kontrafaktiska tillståndet.

Ju mindre dessa kriterier är uppfyllda, desto större är mervärdet av att välja ett upplägg som kan isolera effekterna, och omvänt. Om det är möjligt att välja mellan olika upplägg måste dessa mervärden vägas mot den merkostnad de innebär.

3 Vilken grad av kvalitet och kostnad är de olika alternativen för informationsinhämtning associerade med?

Centralt är att finna information av god kvalitet, som väl beskriver de tänkta effekterna. Som vi såg i 9.2 kan detta vara allt ifrån i princip gratis (kategori 1) till fördyrande (kategori 3) eller mycket kostsamt och kanske olämpligt (kategori 2). Nyttan med att välja en metod för informationsinhämtning som ger effektmått av högre kvalitet, måste alltså vägas mot dess merkostnader (i termer av såväl ekonomiska som etiska kostnader).

Dessa tre kriterier rekommenderas alltså som utgångspunkt för val av insatser som ska bli föremål för effektmätning, i de fall en eventuell effektmätning är associerad med en betydande resursåtgång. Som framgår rekommenderar vi att det strategiska och lärande perspektivet står i fokus. Observera att det är särskilt angeläget att kombinera en resurs-

krävande kvantitativ mätning med god användning av kvalitativ metod (kap. 10) i syfte att utveckla kunskapen om varför, eller varför inte, en insats har lett till goda effekter.

11.3 Vad gör man med resultatet av en effektmätning?

Vi vill avsluta denna del som vi började den, genom att poängtera att det inte finns någon motsättning mellan sofistikerad kvantitativ metod och en kvalitativ metod. Dessa ansatser berikar varandra och bör, om möjligt, användas tillsammans.

Vi har som utgångspunkt för såväl den kvantitativa som den kvalitativa analysen föreslagit den utökade verksamhetslogiken. På samma sätt bör resultatet av mätningen knytas an till verksamhetslogiken. Nedan ges ett par fiktiva exempel med förslag till hur slutsatser kan dras, med hjälp av verksamhetslogiken.

1. Den kvantitativa mätningen har påvisat endast små effekter av insatsen. Utifrån verksamhetslogiken har vi identifierat ett par länkar i kedjan som är svaga eller bristfälliga, trots att utvärderingen av insatsen pekar på att dessa var nödvändiga för att insatsen skulle ge goda effekter.

Orsaken till de svaga länkarna kan vara flerfaldig. Genom att ha identifierat de ramfaktorer som är särskilt centrala för insatsen, skulle resultatet av effektmätningen kunna bli att särskilda ramfaktorer ses över och förstärks. Det kan handla om organisationskultur, IT-stöd, kompetens, m.m. Det kan också vara så att verksamhetslogiken som helhet – eller enstaka länkar i kedjan – och de antaganden som hör därtill, med rätta ifrågasätts. Det kan finnas såväl teoretiska som empiriska skäl utifrån andra studier att diskutera giltigheten i dessa.

2. Den kvantitativa mätningen visar att insatsen gett mycket goda effekter. Genom en analys av hållbarheten av verksamhetslogiken har man identifierat att det finns goda skäl att förvänta sig goda effekter utifrån forskning inom relevanta områden samt erfarenheter dragna inom t.ex. andra skatteadministrationer. Vidare har de ramfaktorer som utgjorde risker för en lyckad implementering av insatsen, visat sig hålla måttet. Resultatet av denna effektmätning skulle kunna vara att insatsen utvidgas, t.ex. i form av mera resurser eller mer omfattande implementering (t.ex. att fler företagsgrupper med liknande problematik blir föremål för insatsen), förutsatt att resultaten kan förmodas vara giltiga även för dessa.

Sammanfattningsvis presenterar del tre en rad verktyg för hur man kan utvärdera effekterna av Skatteverkets insatser och hur genomförda effektmätningar kan ligga till grund för ett kontinuerligt lärande.

Litteraturförteckning

Referenser:

Australian Taxation Office (ATO), *Literature review – Measuring compliance effectiveness*, 2007

Campbell, Donald T. *Assessing the Impact of Planned Social Change*, Paper 8#, Occasional papers series, Dartmouth College 1976

Delhey, Jan & Newton, Kenneth, *Social Trust: Global Pattern or Nordic Exceptionalism?*, Wissenschaftszentrum Berlin für Sozialforschung (WZB) 2004

Diamond, Alexis, *Reliable estimation of average and quantile causal effects in non-experimental settings*, 2005, Tillgänglig på www.people.fas.harvard.edu/~rtmoore/GMMC/Quantile.pdf

Diamond, Alexis & Sekhon, Jasjeet S. *Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies*, 2006. Tillgänglig på <http://sekhon.berkeley.edu/papers/GenMatch.pdf>

ESV 2001:16, *Verksamhetslogik – ger överblick, åstadkommer delaktighet och skapar förståelse*

ESV 2006:7, *Måluppfyllelseanalys – Hur måluppfyllelse, effekter och effektivitet kan undersökas och rapporteras*

ESV 2006:8, *Effektutvärdering – Att välja upplägg*

ESV 2007:23, *Resultat och styrning i statsförvaltningen*

Frankfort-Nachmias, Chava & Nachmias, David, *Research Methods in the Social Sciences*, 5 uppl. London 1996

Gujarati, N. Damodar, *Basic econometrics*, Fourth international edition, Mc Graw Hill, 2003

Heckman, James, *Sample selection bias as a specification error*. *Econometrica*, 1979:47, 153-161.

HM Revenue and Customs 2008, KAI Evaluation guidance series, *Number 3: Evaluation Methodologies* (draft version)

IFAU-rapport 2008:1, *Effekter av yrkesinriktad arbetsmarknadsutbildning för deltagare under perioden 2002–04*

Independent Evaluation Group (IEG), World Bank 2007, *Impact Evaluation – the Experience of the Independent Evaluation Group of the World Bank*

King, Gary m.fl., *A 'politically robust' experimental design for public policy evaluation, with application to the Mexican universal health insurance program*, *Journal of policy analysis and management*, 26(3):479-509, 2007

Newbold, Paul m.fl., *Statistics for business and economics*, fifth edition, Prentice hall, 2003

Newcomer, Kathryn E. (ed.), *Using Performance Measurement to Improve Public and Nonprofit Programs*, *New Directions for Evaluation* no. 75, San Fransisco 1997

Nordisk benchmarking, *Temaafsnit: Effektmål og effektmålinger*, 2007

OECD, Forum on Tax Administration: Compliance sub-group. *Measuring Taxpayers' Compliance: A practical guide based on leading revenue body experience* [incomplete draft report], November 2007

Perrin, Burt, *Effective use and misuse of performance measurement*, *American Journal of Evaluation*, vol. 19, no. 3, 1998, s. 367-379

Pollitt, Christopher, *The Essential Public Manager*, 2003, Open University Press, refererad i SOU 2007:75

Rieper, Olaf & Foss Hansen, Hanne, *Metodedebatten om evidens*, AKF Forlaget 2007

RSV Rapport 1997:1, *Resultatdialogen mellan RSV och SKM* (Resumé)

RSV Rapport 2001:1, *Om gratislotter och preventiv effekt*

Sehkon, Jasjett S., *The Neyman-Rubin model of causal inference and estimation via matching methods*, kommande i *The Oxford handbook of political methodology*

SKV Rapport 2004:9, *Grundläggande uppföljningsinformation* (Guppi)

SKV Rapport 2006:3, *Krånglig moms – en företagsbroms?*

SKV Rapport 2007:1, *Svartköp och svartjobb i Sverige*

SKV Rapport 2008:1, *Skattefelskartan i Sverige*

SOU 2007:75, *Att styra staten – regeringens styrning av sin förvaltning*, Betänkande av Styretredningen

Tarschys, Daniel, ”*Huru skall Statsverket granskas?*”, Ds 2002:58, refererad i SOU 2007:75

Todd, Petra E. *A practical guide for implementing Matching estimators*, 1999, tillgänglig på athena.sas.upenn.edu/~petra/papers/ (2008-01-08)

Todd, Petra E., *Matching estimators*, 2006, tillgänglig på athena.sas.upenn.edu/~petra/research/ (2008-01-08)

Verva 2007:12, *Styrning för ökad effektivitet*

Vedung, Evert, *Utvärdering i politik och förvaltning*, Studentlitteratur 1998

Interna Skatteverkspromemorior:

Danielsson, Joacim & Voxberg, Andreas, *Informationsinsats inför KU-avstämningen 2007 med resultat av utförd effektmätning*, 2007

Danielsson, Joacim & Franzon, Henrik, *Effektmätning av införandet av personalliggare och Skatteverkets besökskontroller* (arbetsmaterial), 2008

Voxberg, Andreas, Danielsson, Joacim & Blomkvist, Jan, ”*Cash is king*” – *pilotstudie för ett effektivare borgenärsarbete*, 2007

Bilaga. En fördjupning i metoden för effektmätning

Slumpmässigt kontrollerade experiment

Vi har redan omnämnt champion/challenger-metoden som en tilltalande utgångspunkt vid användning av slumpmässigt kontrollerade experiment för kunskapsuppbyggnad och strategiutveckling.

I det följande kommer vi att närmare presentera några ytterligare exempel på slumpmässiga upplägg som kan vara förenliga med en effektiv verksamhet.

En möjlighet är att bland t.ex. de bedömda högriskföretagen slumpmässigt fördela företagen till insats respektive icke-insats. En annan snarlik lösning är att göra stratifierade slumpmässiga urval, där sannolikheten är stor – men inte 1 – för att högriskföretag ska bli föremål för insats, och liten – men inte 0 – för att lågriskföretag ska bli föremål för insats. Ytterligare en liknande lösning är att slumpmässigt fördela tidpunkten för insats inom den bedömda högriskgruppen, så att ännu icke kontrollerade företag utgör en referensgrupp till de som blivit kontrollerade. Exempelvis kan urval för moms kontroll fördelas slumpmässigt över ett antal redovisningsperioder. Alla företag i högriskgruppen kontrolleras alltså, men inte alla direkt, utan slumpmässigt fördelade i tiden.⁵⁸

Sammanfattningsvis har vi identifierat tre effektiva experimentella upplägg:

- Slumpmässig fördelning av bedömda högriskföretag till insats respektive inte insats
- Fördelning genom stratifierade slumpmässiga urval från hela populationen (där högriskföretag har stor sannolikhet att bli föremål för insats, och lågriskföretag liten)
- Fördela tidpunkten för insats av högriskföretagen slumpmässigt i tiden.

De två första uppläggen är alltid möjliga att genomföra så länge det går att identifiera vad vi här kallat en högriskgrupp. Det sista upplägget förutsätter att effekten av insatsen inträder i nära anslutning till tidpunkten för insatsen samt att insatsen har en viss utbredning i tiden.

Ett potentiellt problem vid slumpmässiga studier är att det kan uppstå goda skäl att genomföra t.ex. en kontroll för ett företag som fördelats till referensgruppen. Vid en stor kontrollinsats kan det också vara svårt att implementera ett experimentellt upplägg, då ett stort antal medarbetare är delaktiga i insatsen. Vid icke obetydliga avvikelser kan *hela* experimentupplägget omkullkastas och följden blir att det är osäkert om estimaten återspeglar den sanna effekten av insatsen.

I ett av de största slumpmässiga experimenten som genomförts i världen tillämpas en metod som syftar till att vara robust mot ovanstående risker.⁵⁹ Upplägget, som omnämns *randomized cluster matched pair design*, är egentligen inte komplicerat. Den slumpmässiga fördelningen till insats görs i två steg. Först tillämpas en matchningsmetod (se nedan för exempel) för att erhålla ett stort antal parvis liknande objekt. Därefter sker slumpmässig fördelning *per par* till antingen insats, eller inte insats. Vid eventuella avvikelser vid den faktiska implementeringen (så att t.ex. *båda* objekten i ett antal par blir föremål för insatsen) blir inte estimaten av insats-effekten lidande. De par för vilka avvikelser skett, exkluderas helt enkelt ur studien. Vi vill rekommendera denna ansats då större insatser planeras bli föremål för ett experimentellt upplägg. Ansatsen kan också kombineras med metod 1 och 3 ovan.

⁵⁸ Identifiering av högriskföretag kan göras utifrån s.k. affärsregler (t.ex. Skattverkets gemensamma urvalspaket) eller operativt. Den slumpmässiga fördelningen, till insats eller inte respektive till tidpunkt för insats, måste dock göras samlat för hela den identifierade högriskgruppen.

⁵⁹ King m.fl. 2007

Att estimerar effekter vid icke-experimentella studier

Frågorna *om* och *när* det är möjligt att i efterhand finna en god referensgrupp vid icke-experimentella insatser, och i så fall *hur*, har rönt stort intresse inom forskningsvärlden de senaste decennierna och metodutvecklingen är stark.⁶⁰ Särskilt rör sig diskussionerna kring metoder för efterhandsmatchning mellan objekt (personer, företag, etc.) som varit föremål för insats, deltagit i program, etc. och objekt som inte varit föremål för insatsen, etc. Matchningen görs för att efterlikna en experimentell situation. Den syftar till att estimerar kausala effekter av insatsen, och har blivit allt populärare inom en rad discipliner och används av utvärderare inom t.ex. internationella Världsbanken och svenska Institutet för arbetsmarknadspolitisk utvärdering (IFAU).⁶¹

Introduktion till matchningsmetoder

I matchningen ligger viljan att finna objekt som liknar de som blivit föremål för insatsen i alla relevanta avseenden, förutom att de just inte blivit föremål för insatsen.⁶² Genom att jämföra dessa grupper vill man kunna säga vilken effekt insatsen har gett på de som blivit föremål för den. Gemensamt för dessa matchningsmetoder är att de alla använder bakgrundsinformation om (idealt sett) samtliga objekt i populationen, inte minst genom att beakta de principer som varit vägledande för urvalet. Utifrån denna information matchas objekten.

Den information som ligger till grund för urvalet kan i allmänhet fördelas i kategorierna

- icke-observerbar information
- observerbar, men inte tillgänglig information, samt
- observerbar och tillgänglig information.

Observera att endast information som är observerbar *och* tillgänglig kan användas.

Om t.ex. beslutet att delta i en introduktionsutbildning i hög grad beror på motivation och utbildningsnivå, snarare än den information som myndigheten har om företagaren (företagarens företagsform, bransch, komplexitetsgraden i redovisningen av verksamheten, etc.), kommer matchningen att bli missvisande. Vid myndighetens urval av företag som revideras ett givet år, bygger urvalet på såväl information som är tillgänglig för samtliga företagare som information som inhämtas i det enskilda fallet. Om informationen som ligger till grund för urvalet primärt tillhör den första kategorin (information av god kvalitet finns för hela populationen) är möjligheterna goda att finna en god referensgrupp till de reviderade företagen, särskilt som principerna bakom urvalet är kända.

En viktig gemensam förutsättning för att kunna använda dessa metoder är att sannolikheten för att ett objekt skulle bli föremål för insatsen var större än 0 och mindre än 1 för samtliga objekt i populationen. Detta gäller oavsett om beslutet om att bli föremål för insatsen/programmet var företagarens eller myndighetens. Om insatsen däremot har riktats genom ett urval med en s.k. affärsregel⁶³ (såsom Skatteverkets urval för utökad kontroll) är situationen en annan. Från statistiskt perspektiv är det då frågan om en *totalundersökning*, och dessa matchningsmetoder är inte tillämpliga.

⁶⁰ En av pionjärerna på detta område var James Heckman (1979), som år 2000 erhöll Sveriges Riksbanks pris i ekonomisk vetenskap till Alfred Nobels minne.

⁶¹ Se exempelvis Diamond och Sekhon (2006) för exempel och referenser, vidare Världsbankens oberoende utvärderingsgrupp, IEG. Se också flera rapporter från IFAU, t.ex. IFAU 2008:1.

⁶² Matchning kan göras såväl parvis som gruppvis.

⁶³ En affärsregel är ett logiskt villkor. Endast den skattebetalare som uppfyller villkoret blir föremål för den aktuella insatsen.

Kvalitativ eftermatchning

Vi har redan nämnt den kvalitativa matchningen (se avsnitt 8.2.4). Detta är en vanligt förekommande metod inom samhällsvetenskaper och närliggande vetenskaper. Huvudprincipen är den att man matchar utifrån välgrundad teori.

Full/perfekt matchning

Perfekt matchning till varje objekt ur gruppen som blev föremål för insatsen – d.v.s. att finna referensobjekt som för varje relevant faktor har identiska värden som insatsobjektet – är sålän möjligt, särskilt inte om det handlar om kontinuerliga faktorer. Valet av vilka faktorer bland dessa som ska användas, bygger på en teoretisk modell, och frågan är också vilken vikt respektive faktor ska ha i matchningsprocessen. Låt oss t.ex. säga att vi vill finna en matchning till ett 3-årigt företag inom livsmedelsbranschen med en årlig omsättning om 1 miljon kronor. Vilken kandidat är den bästa; ett 3-årigt företag inom livsmedelsbranschen med en omsättning om 2 miljoner kronor, eller ett 5-årigt företag inom livsmedelsbranschen med en omsättning om 1 miljon kronor?

Propensity score matchning

Ett sätt att hantera osäkerheten av val av variabler och deras vikt i matchningen är en empirisk modell (vanligen en statistisk regressionsmodell av typen logit eller probit). Modellen avser att förklara vilka som blev föremål för insatsen och inte, med hjälp av den bakgrundsinformation som finns tillgänglig. Viktigt är att som bakgrundsinformation (kovariater) använda information som inte påverkas av själva insatsen, särskilt kan egenskaper som gällde *före* insatsen påbörjades vara värdefull att använda.⁶⁴ Ett centralt antagande är att vi kan observera och använda de egenskaper som både påverkar sannolikheten att bli föremål för en insats och effekten av densamma.⁶⁵ Allra mest lämplig är också denna metod att använda då denna information finns lättillgänglig i registerdata, och inte kräver extra informationsinsamling, t.ex. genom en enkätundersökning.

Resultatet av modellering är ett predikerat värde på sannolikheten för att objektet ingick/ingår i studien.⁶⁶ Matchningen sker med avseende på denna s.k. propensity score.⁶⁷ Matchningen sker parvis⁶⁸ och eftersom man kan vänta sig högre sannolikhetsvärden för objekt i experimentgruppen matchas objekten från experiment- och referensgruppen endast inom det gemensamma sannolikhetsintervallet. Medelvärde av effekterna av varje parvis jämförelse är ett estimat på den genomsnittliga effekten av insatsen.

Denna modellering har använts tämligen frekvent de senaste decennierna, t.ex. av Världsbankens oberoende utvärderingsgrupp.⁶⁹

⁶⁴ Todd 1999

⁶⁵ IFAU 2008:1

⁶⁶ Eg. den betingade sannolikheten, givet bakgrundsvariablerna som inkluderas i modellen;

$p(x) = P(T = 1 | X = x)$, där T är en binär variabel "treatment"-variabel och X bakgrundsvariablerna.

⁶⁷ Matchning kan ske enbart utifrån propensity score, eller tillsammans med flera matchningsvariabler.

⁶⁸ Det finns även metoder som bygger på Propensity score matching vilka använder viktade värden från flera referensobjekt, eller t.o.m. samtliga i referensgruppen.

⁶⁹ Världsbankens oberoende utvärderingsgrupp, IEG, 2007

Multivariat matchning

En annan vanlig matchningsmetod bygger på avstånd i en geometrisk rymd, utifrån Mahalonbis avstånd (*eng.* Mahalanobis distance). Denna metod tar hänsyn till hur de statistiska fördelningarna ser ut för de variabler som bakgrundsinformation utgör, och hur dessa variabler är korrelerade med varandra.⁷⁰ Låt oss exempelvis säga att vi använder tre variabler för denna matchning. Då kommer våra objekt att fördela sig i ett 3-dimensionellt rum. Matchning enligt denna metod sker alltså med avseende på avståndet, som sagt med hänsyn till hur dessa tre variabler är korrelerade, mellan punkterna i detta 3-dimensionella rum. Om vi har ett godtyckligt antal variabler, N, som utgångspunkt för matchningen identifieras på motsvarande sätt punkter som ligger nära varandra i det N-dimensionella som dessa variabler utgör.

Genetisk matchning

Sekhon och Diamond (2007) och Diamond (2005) föreslår en matchningsmetod som kan beskrivas som en generalisering av ovanstående metoder, s.k. Genetisk matchning. Bakgrunden är, enligt forskarna, att ovanstående metoder – utifrån såväl dess teoretiska egenskaper som empiriska resultat – under vissa förutsättningar ger estimat av effekter som inte är tillförlitliga eller robusta för små förändringar av data.⁷¹ Genetisk matchning ger, enligt författarna, alltid bättre matching och gäller under bredare antaganden. En av kovariaterna (inputvariablerna) till modellen är lämpligen en känd eller estimerad propensity score. Metoden kan då betraktas som en generalisering av de ovan nämnda metoderna.⁷² Metoden bygger på den grundläggande idén är att använda en genetisk optimeringsalgoritm för val av vikter för var och en av kovariaterna.⁷³

Rekommendation – val av matchningsmetod

Metodutvecklingen på området är som sagt stark, och det är möjligt att det kan vara aktuellt att ompröva denna rekommendation, men när det gäller användning av matchningsmetoder för att estimerar effekter av omfattande och strategiskt viktiga insatser rekommenderas tillämpning av Genetisk matchning.

Instrumentvariabelmetoden

I många fall är det inte möjligt att använda en matchningsmetod, då den information som låg till grund för urvalet inte är observerbar och/eller tillgänglig. Ibland kan det dock vara möjligt att identifiera ett s.k. instrument, d.v.s. en variabel som uppfyller två villkor; den ska (1) åtminstone delvis förklara vilka som blir föremål för en insats eller inte (vara korrelerad med sannolikheten att bli föremål för insatsen), men (2) inte ha någon betydelse för vilken effekt insatsen får. Man kan säga att man söker ett naturligt experiment. Genom att studera hur utfallet varierar för olika värden för instrumentvariabeln får vi ett estimat av insatsens effekter.

⁷⁰ Viktningen bestäms av en s.k. kovariansmatris. Mahalonbis avstånd definieras som

$$d(X_i, X_j) = \left((X_i - X_j)' S^{-1} (X_i - X_j) \right)^{1/2}, \text{ där } S \text{ är kovariansmatrisen.}$$

⁷¹ Under förutsättning att kovariaterna följer en ellipsoid (*eng.* ellipsoidal) fördelning (*ex.* normalfördelning), har Propensity score matching i normalfallet goda egenskaper. Metoden kan dock, i praktiken, t.o.m. försämrade matchningen, även under dessa förutsättningar. (Sekhon kommande)

⁷² Generaliseringen består i att en viktningsmatris, W, adderas till Mahalonbis avståndsmått;

$$d(X_i, X_j) = \left((X_i - X_j)' (S^{-1/2})' W S^{-1/2} (X_i - X_j) \right)^{1/2}, \text{ där } W \text{ är en positivt definit viktningsmatris, och}$$

$S^{1/2}$ är en Cholesky-faktorisering av S, kovariansmatrisen.

⁷³ Matchningsproceduren finns beskriven och implementerad för programpaketet R, se <http://sekhon.berkeley.edu/matching>.

Giltigheten i en instrumentvariabel går dock inte att pröva empiriskt, utan bedömningen om dess giltighet måste göras utifrån välgrundad teori. I vilken mån det skulle gå att identifiera instrument för användning vid t.ex. Skatteverkets kontroll- och informationsinsatser har ännu inte utretts. Det verkar dock, i allmänhet, vara svårt att identifiera instrumentvariabler av god kvalitet.⁷⁴

⁷⁴ Jfr t.ex. Gujarati 2003, s. 679.