



Statistiska centralbyrån    Statistics Sweden

# Kartläggning av felkällor för bättre aktualitet

*Pär Lindholm*

Rapportserien "**Research and Development – Methodology Reports from Statistics Sweden**" publicerar rapporter med resultat från SCB:s forsknings- och utvecklingsverksamhet. Fokus i rapportserien ligger på metodutveckling för offentlig statistikproduktion. Rapportserien publicerar bidrag från alla avdelningar inom SCB och är öppen för bidrag som behandlar en vid mängd av olika metodologiska problem.

Utgivna publikationer i serien:

2006:1 Quantifying the quality of macroeconomic variables

2006:2 Stochastic population projections for Sweden

2007:1 Jämförelse av røjanderiskmått för tabeller

2007:2 Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator.

# Kartläggning av felkällor för bättre aktualitet

*Pär Lindholm*

# Evaluation of sources of error for improved timeliness

Statistics Sweden  
2007

---

Producent  
*Producer*

SCB, utvecklingsavdelningen  
Statistics Sweden, Research and Development Department  
SE-701 89 ÖREBRO  
+ 46 19 17 60 00

Förfrågningar  
*Inquiries*

Pär Lindholm, +46 19 17 69 27  
par.lindholm@scb.se

Det är tillåtet att kopiera och på annat sätt mångfaldiga innehållet i denna publikation.  
Om du citerar, var god uppge källan på följande sätt:

Källa: SCB, Utveckling och forskning – Metodrapporter från SCB, *Kartläggning av felkällor för bättre aktualiteter*.

It is permitted to copy and reproduce the contents in this publication.

When quoting, please state the source as follows:

Source: Statistics Sweden, Research and Development – Methodology Reports from Statistics Sweden, *Kartläggning av felkällor för bättre aktualiteter*.

Cover Ateljén, SCB

ISSN 1653-7149

URN:NBN:SE:SCB-2007-X103OP0703\_pdf (pdf)

*Denna publikation finns enbart i elektronisk form på [www.scb.se](http://www.scb.se)*

## **Förord**

För effektivt beslutsfattande ställs höga krav på det statistiska beslutsunderlaget. Kvalitet innebär att statistiken mäter det som avses att mätas och att mätningarna har stor precision. Därtill är statistikens aktualitet en viktig aspekt för statistikens kvalitet. I en beslutssituation är det angeläget att använda sig av aktuell information. Om statistiken har en avlägsen referensperiod är risken stor att rådande förhållanden är annorlunda än de som impliceras av statistiken. Tid är en resursfråga för produktion av statistik varför det uppstår ett avvägningsproblem mellan aktualitet och andra kvalitetsaspekter. Författaren till denna rapport illustrerar detta avvägningsproblem genom att analysera effekterna av en tidigareläggning av publiceringen av statistik om industrins varuproduktion. Problemet att tidigarelägga publicering av statistik är generellt och rapporten visar på en strukturerad ansats för jämförelse av statistik publicerade vid olika tillfällen.

Statistiska centralbyrån i oktober 2007

Folke Carlsson

Lars Melin

## **Friskrivningsklausul**

I serien Utveckling och forskning – Metodrapporter från SCB, publicerar SCB resultat från utvecklingsarbete rörande metoder och tekniker för statistikproduktion. Författaren/författarna svarar själva för innehåll och slutsatser.



## Innehåll

*A separate text in English is provided at the end of the publication, on page 45.*

Förord .....	3
<b>Sammanfattning .....</b>	<b>7</b>
<b>1. Inledning .....</b>	<b>9</b>
1.1 Bakgrund och Syfte .....	9
1.2 Kort beskrivning av Industrins Varuproduktion .....	10
1.3 Begrepp och förklaringar .....	11
<b>2. Urval och estimation .....</b>	<b>12</b>
2.1 Urvalsdesign .....	12
2.2 Imputeringar och estimation för totalt industrivärde .....	13
2.2.1 Uppdelning på delpopulationer .....	14
2.2.2 Uppdelning på felkällor .....	15
<b>3. Informationsläget vid publiceringstillfällena .....</b>	<b>17</b>
3.1 Bortfall .....	17
3.2 Datakällor .....	18
<b>4. Jämförelse mellan publiceringstillfälle 1 och 2 .....</b>	<b>20</b>
4.1 Jämförelse av skattningen mellan delpopulationer .....	20
4.2 Skattningar av effekten från olika felkällor .....	21
4.3 Analys av datakällorna för delpopulation 4 .....	24
4.4 Sammanfattning av resultaten från jämförelsen mellan publiceringstillfälle 1 och 2 .....	25
<b>5. Korrigering av imputeringsmetod .....</b>	<b>27</b>
<b>6. Diskussion .....</b>	<b>30</b>
<b>Referenser .....</b>	<b>32</b>

---

<b>Bilaga 1</b> .....	<b>33</b>
Manuell imputering (1) .....	33
Företagens Ekonomi (2) .....	34
Skattning av totalt industrivärde framskrivet från IVP år t-1, (3) .....	34
Skattning av industrivärde utifrån SRU- och FEK-uppgifter (4) .....	35
Skattning utifrån SRU- och FEK-uppgifter år t (4b) .....	35
Skattning utifrån SRU- och FEK-uppgifter år t-1 (4a) .....	36
Omsättningsuppgifter i momsdeklarationen från Skatteverket (5) .....	37
<b>Bilaga 2</b> .....	<b>38</b>
Tabellsammanställning .....	38
<b>In English</b> .....	<b>45</b>
Summary .....	45



# Sammanfattning

De fem huvudkomponenterna i kvalitetsbegreppen för officiell statistik är innehåll, aktualitet, tillförlitlighet, jämförbarhet och tillgänglighet. Denna rapport sätter fokus på begreppet aktualitet i en av SCB:s undersökningar, Industrins Varuproduktion.

Undersökningen publicerar nationell statistik nio månader efter referensperiodens slut. I denna studie studeras vilka effekter en tidigareläggning av publiceringen på tre månader har på skattningar av den totala industriverksamheten i undersökningen. I studien görs en uppdelning på några av undersökningens viktigaste felkällor – nämligen granskning, bortfall och imputering. Effekten från respektive felkälla skattas.

Resultaten från studien visar att en tidigareläggning av publiceringen har effekter på skattningarna och att det är imputerings-effekten som är av betydande storlek. Metoderna för att imputera värden visar sig vara känslig för hur informationsläget ser ut vid skattningstillfället. En kritisk grupp av företag pekas ut som sårbara för val av metod. För denna grupp av företag prövas en alternativ imputeringsansats. Den ansatsen använder en mer aktuell datakälla som hjälpinformation för imputering av värden. Den nya ansatsen reducerar imputerings-effekten för denna kritiska grupp av företag och ger på totalnivå överensstämmande skattningar mellan ny och gammal publiceringstidpunkt.



# 1. Inledning

## 1.1 Bakgrund och Syfte

De fem huvudkomponenterna i kvalitetsbegreppen för officiell statistik är innehåll, aktualitet, tillförlitlighet, jämförbarhet och tillgänglighet. Denna rapport sätter fokus på begreppet aktualitet i en av SCB:s större undersökningar. Kortfattat kan man beskriva statistikens aktualitet med statistikens relation till "nuläget". Ju närmare sin referenstid statistiken publiceras desto mer aktuell eller snabb är statistiken och vanligen mer intressant.

Undersökningen Industrins Varuproduktion publicerar idag officiell nationell statistik nio månader (den sista september) efter referensårets slut. Med önskemål om en snabbare statistik både internt inom SCB och av externa användare så görs i den här utredningen en översyn av möjligheterna att publicera officiell statistik redan sex månader (sista juni) efter referensårets slut.

Statistiken sammanställs och levereras enligt en EU-förordning (PRODCOM) till Eurostat den 30 juni. Det finns således förutsättningar för att även publicera nationellt den 30 juni, dock till en något sämre kvalitet än till den 30 september. Den centrala frågeställningen är givetvis hur mycket kvaliteten försämras av att tidigarelägga publiceringen tre månader och med vilka åtgärder och statistiska metoder som vi kan parera effekten av de faktorer som orsakar försämringar i kvaliteten.

Det totala industrivärdet för de två senaste undersökningsåren, 2004 och 2005, skattas högre i juni än i september. För år 2004 är överskattningen 8,5 miljarder kronor, vilket motsvarar cirka 0,6 procent av det totala industrivärdet. För år 2005 är överskattningen ungefär 2,1 miljarder kronor.

I denna rapport ska vi försöka kartlägga skillnaderna mer i detalj. Vi gör en gruppering av företagen i populationen och studerar skillnader i skattningarna mellan juni och september för dessa grupper. Vi ger en beskrivning av informationsläget vid de olika publiceringstillfällena och vi försöker att isolera och mäta effekten från de viktigaste felkällorna, såsom granskning, bortfall och imputering.

Kartläggningen avgränsas i denna rapport till att enbart omfatta skillnader i det totala industrivärdet. Skillnader på varunivå behandlas inte i denna framställning då rapporten skulle bli alltför omfattande och detaljerad. Att få en överensstämmelse i skattningarna mellan de båda publiceringstillfällena för det totala industrivärdet är ett nödvändigt men inte tillräckligt kriterium för att tidigarelägga publiceringen. En förutsättning är även att överensstämmelsen på varunivån är hög.

Jämförelserna av statistiken mellan publiceringstillfällena i juni och september studeras för de två senaste undersökta åren, 2004 och 2005. Anledningen till att jämförelsen avgränsas till de två senaste årgångarna är att för undersökningsåret 2004 genomfördes stora förbättringar i undersökningens imputeringsystem. Bland annat genererar det systemet processdata vilket har möjliggjort denna analys.

## 1.2 Kort beskrivning av Industrins Varuproduktion<sup>1</sup>

Industrins Varuproduktion (IVP) är en årlig undersökning som syftar till att beskriva den svenska industrins produktion av varor och industriella tjänster (t ex reparationer och lönebearbetning) på en detaljerad nivå. Statistiken redovisas enligt den Kombinerade Nomenklaturen (KN) som består av drygt 10 500 varunummer där ett varunummer kan sägas motsvara en vara. Uppgifterna samlas in från företag där varans produktionsvärde (varans värde vid leverans) och tillhörande volymuppgift (ton, kubikmeter, liter etc.) är centrala variabler för undersökningen. IVP redovisar produktionsvärden och volymuppgifter på cirka 3 900 av dessa varunummer. Stora användare av statistiken är Eurostat och Nationalräkenskaperna men statistiken är även starkt efterfrågad av andra användare såsom branschorganisationer, andra myndigheter och företag.

Statistiken levereras till Eurostat den 30 juni året efter referensåret (d v s det år som undersökningen avser) medan en preliminär version av statistiken redovisas i Sveriges Statistiska Databaser (SSD) först den 30 september. Den definitiva statistiken redovisas i SSD den 31 januari. Skillnader i leveranser till Eurostat och vad som

---

<sup>1</sup> För en mer utförlig beskrivning av Industrins Varuproduktion hänvisas till [www.scb.se](http://www.scb.se).

publiceras nationellt i SSD är att statistiken till Eurostat redovisas på en aggregerad nivå och att den inte omfattar samtliga varuområden som den nationella statistiken kräver. Bland annat ingår inte de industriella tjänsterna i leveransen till Eurostat. Dock samlas dessa uppgifter in löpande i undersökningen.

### **1.3 Begrepp och förklaringar**

I rapporten blandas uttrycken publiceringstillfälle 1 och 2 samt juni och september månads publicering. Publiceringstillfälle 1 är synonymt med juni och publiceringstillfälle 2 är synonymt med september.

## 2. Urval och estimation

### 2.1 Urvalsdesign

Urvalsprocessen i IVP beskrivs översiktligt i denna rapport. Fokus ligger mer på imputerings- och estimationsprocessen.

Målpopulationen (av företag) i IVP avser mängden företag inom tillverkningsindustrin (SNI 10–37) som har haft verksamhet under referensåret. I urvalsdesignen så delas populationen upp i en del som undersöks med hjälp av blankett, d v s en blankettpopulation, och en del som inte undersöks alls, d v s en cutoff-population.

Vilka företag som hamnar i blankettpopulationen styrs av en EU-förordning som ger riktlinjer om att samtliga företag med 20 anställda eller fler inom tillverkningsindustrin omfattas av uppgiftslämnarplikt. IVP gör dock ett avsteg från EU:s riktlinjer och undantar vissa företag från uppgiftslämnarplikt på grund av kostnadsskäl för undersökningen. De företagen ska inte ingå i blankettpopulationen vart annat år. Dessa företag utgör varannan års populationen<sup>2</sup>.

Blankettpopulationen innehåller således alla företag med 20 anställda eller fler, exklusive de företag som i referensåret ingår i varannan års populationen. Blankettpopulationen totalundersöks. För de företag som inte tillhör blankettpopulationen modellberäknas deras varuproduktion.

Underlaget för att approximera målpopulationen av företag utgår från Företagsregistret (FDB) på SCB. Totalt har målpopulationen omfattat cirka 62 000 – 64 000 företag de senaste åren.

---

<sup>2</sup> Den regel som bestämmer vilka företag som klassificeras som ett varannan års företag ser till företagets produktionsvärde och vilka varor som företaget producerar. Anses företaget ha en mindre viktig varuproduktion är företaget aktuellt för att bli ett varannan års företag. Ett ytterligare villkor är att företaget föregående år redovisade uppgifter i undersökningen. Storleksmässigt så omfattar det mellan 750 – 1 100 företag årligen. Ingen mätteknisk studie har genomförts för att studera effekten av denna regel.

## 2.2 Imputeringar och estimation för totalt industrivärde

Det totala industrivärdet av varor och tjänster för tillverkningsindustrin ges av

$$t = \sum_U y_k$$

Där  $U$  är mängden företag inom hela tillverkningsindustrin och  $y_k$  är industrivärdet för företag  $k$ .

Det totala industrivärdet,  $t$ , skattas enligt

$$\hat{t} = \sum_U y_{\bullet k} \quad \text{där } y_{\bullet k} = \begin{cases} y_k & \text{för } k \in r \\ \hat{y}_k & \text{för } k \in U - r \end{cases}$$

Här innehåller  $r$  mängden företag som svarar och  $\hat{y}_k$  är det imputerade värdet för företag  $k$ .

Den metod och den hierarkiska ordningen för källorna för vilka imputeringen av  $y$  baseras på ges nedan. Redogörelsen för hur  $\hat{y}$  explicit beräknas för respektive datakälla samt en övergripande beskrivning av datakällan ges i bilaga 1.

### Datakälla och metod

- 1 Manuell imputering
- 2 Industrivärde från undersökningen Företagens Ekonomi år  $t$
- 3 Framskrivit industrivärde från IVP år  $t-1$
- 4a Framskrivna regressionsskattning av industrivärdet från nettoomsättning enligt Standardiserade Räkenskapsutdraget (SRU) från Skatteverket år  $t-1$  (används för publicering i juni)
- 4b Regressionsskattning av industrivärdet från nettoomsättning enligt Standardiserade Räkenskapsutdraget från Skatteverket år  $t$  (används för publicering i september, ersätter då källa 4a)
- 5 Industrivärde från nettoomsättning enligt Momsdeklarationen från Skatteverket år  $t$
- 6 Källa saknas, företag  $k$  åsätts värdet 0

Om företag  $k$  finns i källa 1 hämtas industrivärdet från källa 1 annars gå till källa 2 osv.

Den manuella imputeringen i datakälla 1 görs endast för de stora företagen.

### 2.2.1 Uppdelning på delpopulationer

I urvalsfasen gjordes en uppdelning av målpopulationen i 3 delar. I den följande framställningen delas målpopulationen  $U$  upp i fyra disjunkta delpopulationer.

$$U = \{U_1, U_2, U_3, U_4\}$$

Delpopulation 1 består av mängden svarande företag, dvs  $U_1 = r$ .

Delpopulation 2 består av mängden svarsbortfall, dvs  $U_2 = s - r$ .

Delpopulation 3 består av mängden varannan års företag,  $U_3$ .

Delpopulation 4 består av mängden företag under cutoff-gränsen,  $U_4$ .

Vi kan även skriva  $U_1 + U_2 = s$  och  $U_3 + U_4 = U - s$ .

Där  $s$  är mängden företag som ingår i urvalet.

Totalen för delpopulation 1 skattas således enligt

$$\hat{t}_1 = \sum_r y_k = \sum_{U_1} y_{\bullet k}$$

Totalen för delpopulation 2 skattas enligt

$$\hat{t}_2 = \sum_{s-r} \hat{y}_k = \sum_{U_2} y_{\bullet k}$$

Totalen för delpopulation 3 skattas enligt

$$\hat{t}_3 = \sum_{U_3} y_{\bullet k}$$

Totalen för delpopulation 4 skattas enligt

$$\hat{t}_4 = \sum_{U_4} y_{\bullet k}$$

Den skattade summerade totalen för delpopulation 3 och 4 kan även skrivas som

$$\hat{t}_3 + \hat{t}_4 = \sum_{U-s} y_{\bullet k}$$



Skattningen av det totala industrivärdet  $t$  kan då skrivas som

$$\hat{t} = \hat{t}_1 + \hat{t}_2 + \hat{t}_3 + \hat{t}_4.$$

Vi introducerar två publiceringstillfällen,  $p1$  (juni) och  $p2$  (september). Då kan beteckning  $p1$  och  $p2$  indexeras på  $U, U_1, U_2, U_3, U_4, s, r, \hat{t}, \hat{t}_1, \hat{t}_2, \hat{t}_3, \hat{t}_4$  för att beskriva respektive publiceringstillfälle.

T ex kan det skattade totala industrivärdet skrivas som

$$\hat{t}_{p1} = \hat{t}_{1p1} + \hat{t}_{2p1} + \hat{t}_{3p1} + \hat{t}_{4p1} \text{ för publiceringstillfälle 1 och}$$

$$\hat{t}_{p2} = \hat{t}_{1p2} + \hat{t}_{2p2} + \hat{t}_{3p2} + \hat{t}_{4p2} \text{ för publiceringstillfälle 2.}$$

Den skattade differensen för totalen mellan publiceringstillfälle 1 och 2 ges då av

$$\hat{d} = \hat{t}_{p1} - \hat{t}_{p2}.$$

Differensen kan på motsvarande sätt skattas för respektive delpopulation. Delpopulation 3 och 4 (företag med urvalsannolikhet 0) består av samma mängd företag vid publiceringstillfälle 1 och 2 under antagandet om att målpopulationen inte förändras mellan publiceringstillfällena. Delpopulation 1 och 2 förändras, där delpopulation 1 växer i samma utsträckning som delpopulation 2 krymper mellan publiceringstillfällena.

## 2.2.2 Uppdelning på felkällor

Det finns huvudsakligen tre felkällor vi är intresserade att mäta effekten av, nämligen granskning, bortfall och imputering.

Granskningseffekten definieras som differensen av industrivärdet för de företag som tillhör gruppen svarande vid både publiceringstillfälle 1 och 2.

Granskningseffekten skattas av

$$\hat{d}_{11} = \sum_{U_{11}} (y_{\bullet,k,p2} - y_{\bullet,k,p1}) \quad \text{där } U_{11} = U_{1p1} \cap U_{1p2}$$

Sätt  $\hat{d}_{Gr} = \hat{d}_{11}$ .

Bortfallseffekten är den summerade skillnaden i skattat värde och svarat värde för de företag som inte svarat vid publiceringstillfälle 1 men som inkommit med uppgifter till tillfälle 2. Denna effekt kan även ses som en imputeringseffekt men för att särskilja bortfallet

från de ej urvalsundersökta företagen benämns denna som bortfallseffekt.

Bortfallseffekten skattas av

$$\hat{d}_{21} = \sum_{U_{21}} (y_{\bullet k, p2} - y_{\bullet k, p1}) \quad \text{där } U_{21} = U_{2p1} \cap U_{1p2}$$

$$\text{Sätt } \hat{d}_{Bf} = \hat{d}_{21}.$$

Imputerings effekten är den effekt som inte är gransknings- eller bortfallseffekt. Det är en differens som vi själva har introducerat i skattningen beroende på att informationsläget vid publiceringstillfällena är olika. Det handlar om företag som varken har inkommit med uppgifter vid publiceringstillfälle 1 eller 2.

Imputerings effekten skattas av

$$\hat{d}_{22} = \sum_{U_{22}} (y_{\bullet k, p2} - y_{\bullet k, p1}) \quad \text{där } U_{22} = U_{2p1} \cap U_{2p2}$$

$$\hat{d}_{33} = \sum_{U_{33}} (y_{\bullet k, p2} - y_{\bullet k, p1}) \quad \text{där } U_{33} = U_{3p1} \cap U_{3p2}$$

$$\hat{d}_{44} = \sum_{U_{44}} (y_{\bullet k, p2} - y_{\bullet k, p1}) \quad \text{där } U_{44} = U_{4p1} \cap U_{4p2}$$

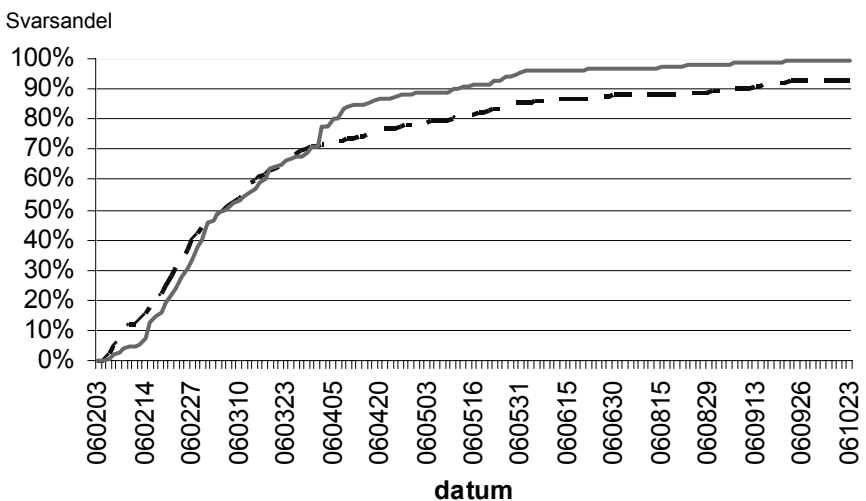
$$\text{Sätt } \hat{d}_{Imp} = \hat{d}_{22} + \hat{d}_{33} + \hat{d}_{44}.$$

## 3. Informationsläget vid publiceringstillfällena

### 3.1 Bortfall

En förutsättning för att tidigarelägga publiceringen i SSD är att bortfallet för de urvalsundersökta företagen inte är för högt i juni. Framförallt är det viktigt att de stora företagen inkommer med uppgifter. I figur 1 nedan visas den kumulativa vägd svarsfrekvensen samt den kumulativa ovägd svarsfrekvensen över insamlingsperioden för år 2005. Den vägd svarsfrekvensen är vägd efter företagets industrivärde och den ovägd svarsfrekvensen beräknas utifrån att alla företag har samma vikt.

**Figur 1**  
**Ovägd och vägd svarsandel, 2005**



Heldragna linjen anger vägd och streckad linje anger ovägd svarsandel.

I figur 1 kan vi utläsa att i mitten på juni så är den vägd svarsfrekvensen cirka 96 procent. Tre månader senare (inför publiceringen i september) så har den vägd svarsfrekvensen ökat till 99 procent. Den vägd svarsfrekvensen är hög redan i juni.

Marginal effekten med att ha en tre månader längre insamlingsperiod är 3 procentenheter.

## 3.2 Datakällor

I avsnitt 2 redogjordes för källordningen för att imputera industrivärdet för ett företag då företaget inte har svarat. Om vi utökar datakällorna i avsnitt 2 till att även omfatta ett värde som inkommit i undersökningen fås ett företags industrivärde enligt följande ordning.

### Datakälla och metod

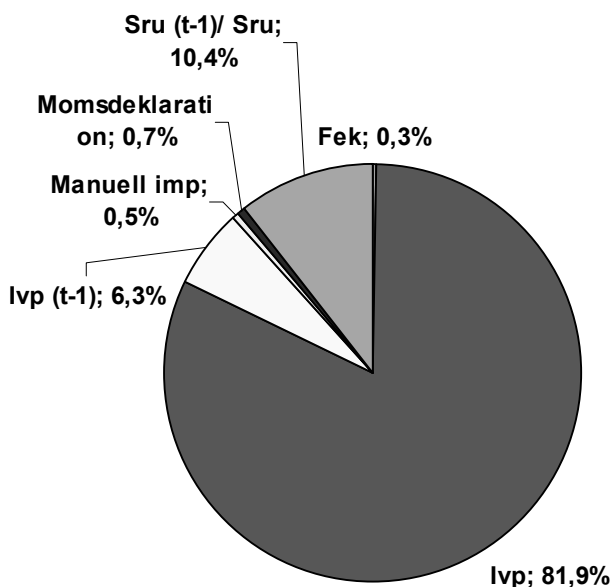
- 1 Industrivärde från IVP år  $t$
- 2 Manuell imputering
- 3 Industrivärde från undersökningen Företagens Ekonomi år  $t$
- 4 Framskrivit industrivärde från IVP år  $t-1$
- 5a Framskrivna regressionsskattning av industrivärdet från nettoomsättning enligt Standardiserade Räkenskapsutdraget (SRU) från Skatteverket år  $t-1$  (används för publicering i juni)
- 5b Regressionsskattning av industrivärdet från nettoomsättning enligt Standardiserade Räkenskapsutdraget från Skatteverket år  $t$  (används för publicering i september, ersätter då källa 5a)
- 6 Industrivärde från nettoomsättning enligt Momsdeklarationen från Skatteverket år  $t$
- 7 Källa saknas, företag  $k$  åsätts värdet 0

En skillnad i källorna mellan publiceringstillfälle 1 och 2 är att de svarande företagen i IVP är färre i juni. Dock såg vi i avsnitt 3.1 att det värdemässigt endast motsvarar tre procentenheter jämfört med september.

Även de svarande företagen i undersökningen Företagens Ekonomi är färre i juni än i september men deras ökade tillskott ger inget stort bidrag till imputeringsunderlaget. En större skillnad i källorna är att årets Standardiserade Räkenskapsutdrag från Skatteverket (se datakälla 5) inte finns tillgänglig vid publiceringen i juni. Istället baseras industriskattningen på föregående års räkenskapsutdrag som sedan framskrivs med branschens utvecklingstal för referensåret (se bilaga 2).

I figur 2 visas fördelningen mellan datakällornas inbördes bidrag till totalskattningen av industriproduktionen vid publiceringstillfälle 1 år 2005.

**Figur 2**  
**Underlag för beräkningar i IVP, andel per källa år 2005**



Av figur 2 framgår att bidraget, beräknat på industrivärden, från de svarande företagen (källa IVP) är ungefär 82 procent i juni. Det procenttalet kan tolkas som den faktiska täckningsgraden i undersökningen. SRU (t-1) bidrar som källa med ungefär 10 procent av totalskattningen i undersökningen. Framförallt är det företagen i delpopulation 4 (företag under cutoff-gränsen) som får SRU som källa. Även källa IVP föregående år, IVP (t-1), ger ett betydande bidrag till underlaget för imputeringen. Det är framförallt bortfallsföretagen och varannan års företagens värde som baseras på den källan.

Till september månad växer IVP:s andel som källa med 3 procentenheter, huvudsakligen på föregående års IVP:s bekostnad. Årets SRU ersätter SRU (t-1) som källa. I övrigt förändras inte figuren nämnvärt.

## 4. Jämförelse mellan publiceringstillfälle 1 och 2

### 4.1 Jämförelse av skattningen mellan delpopulationer

I tabell 1 så redovisas skattningen av den totala industriproduktionen för publiceringstillfälle 1 och 2 för åren 2004 och 2005. Skattningarna redovisas per delpopulation. Även antalet företag redovisas i tabellen.

**Tabell 1**

**Skattat industrivärde per delpopulation. I fetstil anges industrivärdet i tusentals kronor (tkr). Under industrivärdet anges antalet företag per kategori**

	2004		2005	
	P1 (Juni)	P2 (Sept)	P1 (Juni)	P2 (Sept)
$\hat{t}_1$ (svarande)	<b>1 140 696 204</b> 3 467	<b>1 147 932 801</b> 3 521	<b>1 168 769 224</b> 3 070	<b>1 204 479 230</b> 3 250
$\hat{t}_2$ (bortfall)	<b>71 219 209</b> 596	<b>62 204 903</b> 558	<b>56 100 695</b> 572	<b>23 648 437</b> 392
$\hat{t}_3$ (varannan års)	<b>40 506 789</b> 774	<b>40 547 379</b> 774	<b>62 124 282</b> 1 096	<b>62 170 479</b> 1 096
$\hat{t}_4$ (cutoff)	<b>141 340 214</b> 57 275	<b>134 535 524</b> 57 275	<b>143 031 220</b> 58 734	<b>137 591 798</b> 58 734
Undertäckning	<b>18814</b> 16	<b>0</b> 0	<b>0</b> 0	<b>0</b> 0
$\hat{t}$ (total)	<b>1 393 762 416</b> 62 128	<b>1 385 220 608</b> 62 128	<b>1 430 025 420</b> 63 472	<b>1 427 889 943</b> 63 472

För att juni månads skattning av det totala industrivärdet ska betraktas som tillförlitlig ska skillnaden mellan skattningen i september inte vara av betydande storlek. Ur tabellen kan man enkelt beräkna att differensen mellan juni och september är 8 542 mnkr för år 2004 och 2 135 mnkr för år 2005.

Varannan års företagen och företagen under cutoff-gränsen borde ha överensstämmande skattningar mellan de båda publiceringstillfällena. För bortfallsföretagen borde industrivärdet ha minskat till september och istället ökat för de svarande företagen.

Det är endast totalen för varannan års företagen,  $\hat{t}_3$ , som skattas överensstämmande. Att inte  $\hat{t}_1$  ökar i samma utsträckning som  $\hat{t}_2$  minskar mellan tillfälle 1 och 2 tyder på att finns en effekt från en eller flera felkällor och att nettoeffekten inte blir noll.

I tabellen framgår tydligt att  $\hat{t}_4$  i juni rejält överskattar värdet i september. Mönstret är tydligt för båda åren och det visar på att det kan finnas ett systematiskt fel i skattningen av företagen i delpopulation 4.

Det finns två tydliga större diskrepanser i tabellen.

1.  $\hat{t}_2$  minskar inte i samma utsträckning som  $\hat{t}_1$  ökar mellan publiceringstillfälle 1 och 2 för år 2005. Till september har skattningen av värdet ökat med 3 258 mnkr för delpopulation 1 och 2.
2.  $\hat{t}_4$  minskar i värde kraftigt mellan publiceringstillfälle 1 och 2 för båda åren.

Vi ska i nästa avsnitt försöka finna förklaringar till ovanstående diskrepanser.

## 4.2 Skattningar av effekten från olika felkällor

I tabell 2 korstabuleras delpopulationerna mellan publiceringstillfälle 1 och 2 och differensen av industrivärdet beräknas för varje cell. På så sätt kan vi isolera och mäta effekten från respektive felkälla. Utöver differensen beräknas även det totala industrivärdet vid den första publiceringen för att se differensen i relation till en total. Underst anges antalet företag i cellen. Samtliga uppgifter anges även i marginalen. Vi har i det följande valt att enbart redovisa korsstabeller för år 2005 i huvudframställningen. En komplett tabellsammansättning för år 2004 finns i bilaga 2.

**Tabell 2****Korstabell över delpopulationer mellan publiceringstillfälle 1 och 2, år 2005. Värdet i cellen är i tusentals kronor (tkr)**

P1 (juni)		P2 (sept)				
		Delpop p2				
Delpop p1	Skattad Parameter	1 (svarande)	2 (bortfall)	3 (varannan års)	4 (cutoff)	Totalt
1 (svarande)	<b>Differens</b>	<b>3 016 895</b>				<b>3 016 895</b>
	Total P1	1 168 769 224				1 168 769 224
	Antal företag	3 070				3 070
2 (bortfall)	<b>Differens</b>	<b>-352 819</b>	<b>593 672</b>			<b>240 853</b>
	Total P1	33 045 930	23 054 765			56 100 695
	Antal företag	180	392			572
3 (varannan års)	<b>Differens</b>			<b>46 196</b>		<b>46 196</b>
	Total P1			62 124 282		62 124 282
	Antal företag			1 096		1 096
4 (cutoff)	<b>Differens</b>				<b>-5 439 421</b>	<b>-5 439 421</b>
	Total P1				143 031 220	143 031 220
	Antal företag				58 734	58 734
Totalt	<b>Differens</b>	<b>2 664 076</b>	<b>593 672</b>	<b>46 196</b>	<b>-5 439 421</b>	<b>-2 135 477</b>
	Total P1	1 201 815 154	23 054 765	62 124 282	143 031 220	1 430 025 420
	Antal företag	3 250	392	1 096	58 734	63 472

Av tabell 2 framgår att den totala differensen är 2 135 mnkr som även framgick av tabell 1. Den skattade granskningseffekten,  $\hat{d}_{11}$ , är 3 017 mnkr. Det betyder att för de företag som inkommit till juni så har dessa företags totala industrivärde korrigerats upp med 3 017 mnkr till september i granskningsarbetet. Det bör påpekas att detta är nettoeffekten av granskningen. Det finns företag vars värde från juni som har korrigerats ned. Den skattade bortfallseffekten,  $\hat{d}_{21}$ , är -353 mnkr. Den skattade imputeringsseffekten för bortfallspopulationen,  $\hat{d}_{22}$ , är 594 mnkr. Den stora skillnaden i industrivärdet mellan juni och september för urvals företagen (företag som tillhör mängden  $U_1 \cup U_2$ ) beror således på granskningseffekten. Bortfalls- och imputeringsseffekten är av mindre storlek, både i absoluta och relativa tal.

Det är 34 företag som bidrar till granskningseffekten år 2005. Både positiva och negativa förändringar har gjorts. Det är ett företag som



kraftigt har korrigerats upp och till stor del bidrar till den stora granskningseffekten.

Imputeringseffekten för varannan års företagen,  $\hat{d}_{33}$ , är 46 mnkr vilket är av blygsam storlek. Imputeringseffekten för företag under cutoff-gränsen,  $\hat{d}_{44}$ , är -5 439 mnkr vilket däremot måste betraktas som stort. För år 2004 är imputeringseffekten för denna delpopulation ännu större, nämligen -6 805 mnkr. Se tabell 2 i bilaga 2.

För att enklare se nettoeffekterna från respektive felkälla görs en sammanställning utifrån tabell 2 i nedanstående tabell.

**Tabell 3**  
**Skattade nettoeffekter för åren 2004 och 2005. Värdet i cellerna är i tkr**

Typ av effekt	År	
	2004	2005
Granskningseffekt $\hat{d}_{11}$	92 210	3 016 895
Bortfalleffekt $\hat{d}_{21}$	-1 119 238	-352 819
Imputeringseffekt (summerad) $\hat{d}_{22} + \hat{d}_{33} + \hat{d}_{44}$	-7 421 323	-4 799 553
Övriga effekter	93 457	
<b>Total nettoeffekt</b>	<b>-8 541 808</b>	<b>-2 135 477</b>

Imputeringseffekten är den klart dominerande effekten för 2004. Två effekter är utmärkande för 2005, gransknings- och imputerings-effekten. Felen tar till viss del ut varandra vilket gör att det totala felet blir 2 136 mnkr för 2005. Det motsvarar 0,1 procent av den totala industriproduktionen.

Det är anmärkningsvärt att skillnaden i skattat värde för företag i delpopulation 4 är stor mellan publiceringstillfälle 1 och 2. Felet är stort både i absoluta och relativa tal. Det förefaller vara av systematisk karaktär. Det är inte möjligt att utifrån de tabeller vi presenterat se vad som orsakar denna stora effekt varför vi gör en mer fördjupad analys över datakällorna i avsnitt 4.3 för delpopulation 4.

### 4.3 Analys av datakällorna för delpopulation 4

På motsvarande sätt som vi korsade delpopulationerna mellan tillfälle 1 och 2 så skapar vi en korstabell över datakällorna mellan tillfälle 1 och 2 för delpopulation 4. För varje cell skattas differensen. I cellerna anges även totalt industrivärde och antalet företag och i marginalerna anges summorna.

**Tabell 4**

**Korstabell över källorna för företag i delpopulation 4 mellan publiceringstillfälle 1 och 2, år 2005. Värdena i cellerna är i tkr**

P1 (juni)		P2 (sept)				
		Källa p2				
Källa p1	Skattad Parameter	Manuell imputering	Sru (t)	Moms	Källa saknas	Totalt
Manuell imputering	<b>Differens</b>	<b>0</b>				<b>0</b>
	Total P1	1 364 608				1 364 608
	Antal företag	65				65
Sru (t-1)	<b>Differens</b>	<b>-160 343</b>	<b>869 049</b>	<b>-2 633 729</b>	<b>-3 417 481</b>	<b>-5 342 503</b>
	Total P1	160 343	124 659 204	4 083 284	3 417 481	132 320 312
	Antal företag	4	44 490	745	3 112	48 351
Moms	<b>Differens</b>	<b>-199 633</b>	<b>-1 054 773</b>	<b>0</b>		<b>-1 254 406</b>
	Total P1	199 633	8 255 503	891 164		9 346 300
	Antal företag	3	2 589	364		2 956
Källa saknas	<b>Differens</b>		<b>1 157 488</b>		<b>0</b>	<b>1 157 488</b>
	Total P1		0		0	0
	Antal företag		3 575		3 787	7 362
Totalt	<b>Differens</b>	<b>-359 976</b>	<b>971 764</b>	<b>-2 633 729</b>	<b>-3 417 481</b>	<b>-5 439 421</b>
	Total P1	1 724 584	132 914 707	4 974 448	3 417 481	143 031 220
	Antal företag	72	50 654	1 109	6 899	58 734

Vi har tidigare konstaterat att differensen mellan de olika publiceringstillfällena är stor för företag under cutoff-gränsen. Även i tabell 4 visas att den totala differensen uppgår till 5 439 mnkr, vilket motsvarar cirka 4 procent av det totala industrivärdet för denna delpopulation. För år 2004 är denna differens ännu större.

De företag som har de lägst prioriterade datakällorna, källa "moms" eller "källa saknas" vid publiceringstillfälle 1 är företag som saknade industriverksamhet föregående år enligt de högre prioriterade källorna. Företag i kategorin "källa saknas" vid publiceringstillfälle 2 är företag som saknar industriverksamhet aktuellt år. Ur tabellen kan man utläsa att det fanns 3 112 företag som hade industriverk-

samhet år 2004 men inte år 2005. Deras skattade industrivärde uppgick till 3 417 mnkr för år 2005. Det fanns 745 företag som hade industriverksamhet år 2004 baserat på det årets SRU men som 2005 fått ett imputerat värde baserat på momsdeklarationen. Båda dessa kategorier utgörs till stor del av företag som har avslutat eller trappat ned sin verksamhet eller bytt organisationsnummer varför det sker en kraftig överskattning av dessa företags industrivärde vid publiceringstillfälle 1.

På motsvarande sätt ser vi att det fanns 3 575 företag som har källa SRU 2005 men saknade SRU för år 2004. Denna kategori företag kan ses som nystartad verksamhet eller så har företaget bytt organisationsnummer. Dock tar inte effekten mellan avslutade och nystartade företag ut varandra. Vi skattar troligtvis de avslutade företagen och företag som trappat ned sin verksamhet alltför högt när det imputerade värdet baseras på föregående års SRU. Liknande mönster kan ses för produktionsåret 2004. Det verkar finnas ett systematiskt fel som genererar dessa skattningskillnader.

Företag som vid tillfälle 2 får ett imputerat värde baserat på årets SRU, får vid korsning av andra källor vid tillfälle 1 en imputeringskillnad. Om vi studerar korsningen mellan källorna årets SRU vid tillfälle 2 och SRU (t-1) vid tillfälle 1 så är skillnaden 869 mnkr. Det kan indikera att utvecklingstalen som används för att framskriva föregående års värde är något för låga. Samma mönster kan ses för år 2004. Dock är det relativa felet väldigt litet för båda åren. Denna effekt är överraskande låg. Det fanns misstankar om att utvecklingstalen skulle ge upphov till större skillnader.

För de företag som i juni månad har fått ett imputerat värde baserat på momsdeklarationen och som i september sedan imputerats med årets SRU så är det relativa felet ganska stort (13 procent). Samma stora relativa fel kan inte påvisas för 2004. I absoluta tal är felet inte av betydande storlek.

#### **4.4 Sammanfattning av resultaten från jämförelsen mellan publiceringstillfälle 1 och 2**

Det finns två betydande diskrepanser i statistiken mellan publiceringstillfälle 1 och 2. Den första skillnaden är en granskningseffekt och den andra skillnaden är en imputeringsseffekt som beror på att informationsläget för de båda publiceringsomgångarna skiljer sig åt.

Vi har tidigare fört ett resonemang kring att granskningseffekten är mindre känslig för en tidigareläggning av publiceringen. Denna effekt kan kontrolleras genom att tillsätta mer granskningsresurser tidigare i processen.

Imputeringseffekten betraktas som mer allvarlig. Vi kan härleda den stora imputeringseffekten till delpopulation 4 och den förefaller vara systematisk. Vi ska i avsnitt 5 försöka finna förklaringar till denna systematiska effekt och försöka anpassa en metod som reducerar effekten.

## 5. Korrigerig av imputeringsmetod

Vi har på mikronivå studerat den grupp av företag som vi misstänker systematiskt överskattar industrivärdet för referensåret, baserat på föregående års SRU. Det finns ett tydligt mönster, att företag som i populationen är inaktiva, skattas alldeles för högt. Det är företag som har varit aktiva någon gång under referensåret och som därför ingår i populationen. Att inaktiva företag ingår i populationen beror historiskt på att undersökningen vill fånga säsongsbetonad industriverksamhet. Att denna grupp av företag överskattas med föregående års SRU är något som vi tidigare anat. Däremot är effekten betydligt större än befarat. Ett värde från föregående år (som dessutom framskrivs) är inte en bra approximation för årets industrivärde för denna grupp av företag. Finns det en mer tillförlitlig datakälla att ersätta föregående års SRU med?

Den ansats som undersökningen historiskt har valt vid värdering av datakällor är att uppgifter från momsdeklarationen ska användas i en så liten utsträckning som möjligt. Den har ansetts som mindre tillförlitlig och att den egentligen inte behövs som källa då SRU ska vara mer eller mindre heltäckande. Momsdeklarationen började för första gången användas som datakälla i undersökningen referensåret 2004 och då som ett komplement till SRU. Den stora fördelen med momsdeklarationen är att den är aktuell. Redan i mars finns föregående års uppgifter tillgängliga. En alternativ ansats är att använda dessa aktuella uppgifter för att imputera industrivärdet för gruppen inaktiva företag.

Vi prövade denna alternativa ansats och av resultaten framgår tydligt att för denna grupp av företag ger momsdeklarationen som källa en avsevärt bättre approximation för industrivärdet än föregående års SRU.

I nedanstående tabell, som har samma utseende som tabell 4, visas för varje cell skattad differens och skattad korrigerad differens. För beräkningar av den skattade korrigerade differensen har föregående års SRU ersatts av momsdeklarationen som datakälla för de inaktiva företagen i populationen.

**Tabell 5**  
**Korstabell över datakällorna för delpopulation 4 mellan publiceringstillfälle 1 och 2 för år 2005. Värdena i cellerna är i tkr**

P1 (juni)		P2 (sept)					Totalt
		Källa p2					
Källa p1	Skattad Parameter	Manuell imputering	Sru (t)	Moms	Källa saknas		
Manuell imputering	<b>Differens</b>	<b>0</b>					0
	<b>Korrigerad differens</b>	<b>0</b>					0
	Antal företag	65					65
Sru (t-1)	<b>Differens</b>	<b>-160 343</b>	<b>869 049</b>	<b>-2 633 729</b>	<b>-3 417 481</b>		<b>-5 342 503</b>
	<b>Korrigerad differens</b>	<b>-160 343</b>	<b>1 238 055</b>	<b>-466 857</b>	<b>-713 468</b>		<b>-102 613</b>
	Antal företag	4	44 490	745	3 112		48 351
Moms	<b>Differens</b>	<b>-199 633</b>	<b>-1 054 773</b>	<b>0</b>			<b>-1 254 406</b>
	<b>Korrigerad differens</b>	<b>-199 633</b>	<b>-1 054 773</b>	<b>0</b>			<b>-1 254 406</b>
	Antal företag	3	2 589	364			2 956
Källa saknas	<b>Differens</b>		<b>1 157 488</b>		<b>0</b>		<b>1 157 488</b>
	<b>Korrigerad differens</b>		<b>1 157 488</b>		<b>0</b>		<b>1 157 488</b>
	Antal företag		3 575		3 787		7 362
Totalt	<b>Differens</b>	<b>-359 976</b>	<b>971 764</b>	<b>-2 633 729</b>	<b>-3 417 481</b>		<b>-5 439 421</b>
	<b>Korrigerad differens</b>	<b>-359 976</b>	<b>1 340 770</b>	<b>-466 857</b>	<b>-713 468</b>		<b>-199 531</b>
	Antal företag	72	50 654	1 109	6 899		58 734

I tabell 5 jämförs den skattade korrigerade differensen med den ursprungliga skattade differensen från tabell 4. Av tabellen framgår att imputeringseffekten har minskat från -5 439 mnkr till -200 mnkr, vilket endast motsvarar 0,1 procent av det totala industrivärdet för delpopulation 4. Den minskade differensen syns tydligt för källorna "moms" och "källa saknas" för publiceringstillfälle 2. Även för 2004 minskar imputeringseffekten kraftigt efter korrigerig av imputeringsansatsen.

**Tabell 6**  
**Skattat industrivärde efter korrigerad metod samt under antagande om att granskningseffekten är 0. Industrivärdet är i tkr**

Estimator		2004		2005	
		P1 (Juni)	P2 (Sept)	P1 (Juni)	P2 (Sept)
$\hat{t}_{urisp}$	Ursprunglig metod	1 393 762 416	1 385 220 608	1 430 025 420	1 427 889 943
$\hat{t}_{korr}$	Korrigerad metod	1 387 034 712	1 385 220 608	1 424 785 538	1 427 889 943
$\hat{t}_{korr}$ och $\hat{d}_{11} = 0$	Korrigerad metod utan gransknings-effekt	1 387 126 922	1 385 220 608	1 427 802 433	1 427 889 943

Tabell 6 visar hur skattningen av det totala industrivärdet förändras i junipubliceringen för dels den korrigerade ansatsen och dels i kombination under antagandet att granskningseffekten helt kan kontrolleras, d v s effekten är 0. Granskningseffekten kan kontrolleras genom att tillsätta mer personella resurser tidigare i processen. För båda årgångarna fås en god överensstämmelse i skattningarna mellan juni och septemberpubliceringen.

## 6. Diskussion

De företag som ligger under cutoff-gränsen och således inte blankettundersöks står för cirka 10 procent av den totala industriproduktionen mätt i produktionsvärde. Det är denna delpopulation som genererar de stora skillnaderna i skattningarna mellan juni och september månads publicering. I studien har det tydligt framkommit att informationsunderlag har betydelse vad gäller skattningar av ett företags totala industrivärde. Skattningarna blir sårbara för att vi i juni månad saknar uppgifter från årets SRU och tvingas använda en annan datakälla som imputeringsunderlag. Valet av metod och datakälla blir viktiga faktorer för att göra tillförlitliga skattningar för företag som ligger under cutoff-gränsen.

En grupp kritiska företag har pekats ut i studien. Det är företag som är inaktiva vid urvalstillfället men som har haft aktivitet någon gång under året, exempelvis säsongsbetonad industriverksamhet. Att använda föregående års SRU som underlag för att i junipubliceringen skatta det totala industrivärdet för denna grupp är en bristfällig metod. När istället momsdeklaration prövas som datakälla i juni för denna grupp av företag reduceras kraftigt skillnaderna mellan de båda publiceringstillfällena. Effekterna syns tydligt för båda åren 2004 och 2005. Orsaken till att momsdeklarationen fungerar bättre som datakälla är att den är aktuell. Det sker stora förändringar inom den här gruppen av företag mellan två år. Bland de inaktiva företagen så ingår troligtvis inte bara företag med säsongsbetonad verksamhet utan även företag med direkta nedläggningar, avtrappning av verksamhet eller organisationsnummerbyten inryms i denna grupp. Undersökningen bör överväga om den här gruppen av företag fortsättningsvis ska ingå i rampopulationen över företag under cutoff-gränsen.

Ytterligare en effekt men som inte är lika kritisk för tidigareläggning av publicering är granskningseffekten. För år 2005 var denna effekt relativt hög. Detta kan åtgärdas genom att i större omfattning arbeta med granskning av de stora företagen tidigare i produktionsprocessen.

Vi har i rapporten visat att vi kan identifiera och mäta effekter av de för undersökningen mest betydanden felkällorna. Genom att pröva en alternativ ansats har vi lyckats finna en bättre metod som starkt



bidrar till en mer tillförlitlig och mer aktuell statistik. Genom denna metodförändring finns förutsättningar att publicera officiell statistik sex månader efter referensårets utgång, vilket är en uppsnabbning av statistiken med tre månader.

Som vi i inledningen av rapporten betonade så är överensstämmelsen av det totala industrivärdet mellan publiceringstillfället i juni och september ett nödvändigt men inte tillräckligt kriterium för att tidigarelägga publiceringen. Genom den nya imputeringsansatsen har vi erhållit överensstämmande skattningar av det totala industrivärdet mellan de två tidpunkterna. Vi anser därmed att steg 1 är avklarat och att nästa steg blir att på motsvarande sätt kartlägga eventuella skillnader på varunivå mellan juni och september månads publiceringar.

Om den nya metoden implementeras i produktionen för kommande årgångar bör uppföljningar göras genom att jämföra den gamla och nya metoden för att utvärdera metodens effekter. Detta som ett led i det fortsatta arbetet med att förbättra statistikens kvalitet.

# Referenser

SCB. (2006). Tillräcklig kvalitet och kriterier för officiell statistik.

# Bilaga 1

I den här bilagan beskrivs metoderna för att ta fram det imputerade värdet,  $\hat{y}_k$ , för att skatta industrivärdet,  $y$ , för företag  $k$ .

Källordningen är enligt följande.

## Datakälla och metod

- 1 Manuell imputering
- 2 Industrivärde från undersökningen Företagens Ekonomi år  $t$
- 3 Framskrivit industrivärde från IVP år  $t-1$
- 4a Framskrivna regressionsskattning av industrivärdet från nettoomsättning enligt Standardiserade Räkenskapsutdraget (SRU) från Skatteverket år  $t-1$  (används för publicering i juni)
- 4b Regressionsskattning av industrivärdet från nettoomsättning enligt Standardiserade Räkenskapsutdraget från Skatteverket år  $t$  (används för publicering i september, ersätter då källa 4a)
- 5 Industrivärde från nettoomsättning enligt Momsdeklarationen från Skatteverket år  $t$
- 6 Källa saknas, företag  $k$  åsätts värdet 0

## Manuell imputering (1)

För de mera betydande företagen där man efter vissa maskinella kontroller upptäcker att den maskinella imputeringen ger en alltför missvisande bild av företagets industrivärde ersätts värdet med ett manuellt imputerat värde. Den manuella imputeringen görs utifrån historiska uppgifter i IVP, externa datakällor, årsredovisningar samt granskarens företagsspecifika kunskaper.

Vanligtvis är det företag som på något vis är komplexa som imputeras manuellt. Det är företag som t ex har genomgått organisationsförändringar under referensåret eller företag som är uppdelade på flera insamlingsenheter.

## Företagens Ekonomi (2)

Syftet med undersökningen Företagens Ekonomi (FEK) är att belysa näringslivets struktur med avseende på exempelvis lönsamhet, tillväxt, utveckling, finansiering och produktion. I undersökningen ska företagen redovisa detaljerade företagsekonomiska uppgifter, bland annat omsättningsfördelning varav total industriverksamhet är en redovisningspost.

Saknas ett inkommit industrivärde i IVP för ett företag hämtas uppgiften från FEK. Ett problem är att undersökningarna ligger i olika faser vad gäller insamling och granskning. IVP har en betydligt snabbare insamlings- och granskningsprocess än FEK vilket gör att FEK för många företag inte ger någon information till IVP, varken i juni- eller septemberpubliceringen.

Då industrivärdet från FEK används för delpopulation 4 redovisas dessa under källa SRU.

Om det totala industrivärdet i FEK betecknas som  $y_k^*$  för företag  $k$  skattas det totala industrivärdet i IVP för företag  $k$  genom  $\hat{y}_k = y_k^*$ .

## Skattning av totalt industrivärde framskrivet från IVP år t-1, (3)

Det totala industrivärdet för företag  $k$  år  $t$  framskrivs med hjälp av en kvotimputering där kvoten,  $R_g$ , skattas utifrån de svarande företagens utveckling mellan år  $t$  och år  $t-1$  för den branschgrupp,  $g$ , företag  $k$  tillhör.

$$\hat{y}_{k,t} = \hat{R}_g \times y_{k,t-1} \text{ om } k \text{ tillhör branschgrupp } g \quad (1)$$

$$\text{där } \hat{R}_g = \frac{\sum_{r_g} y_{k,t}}{\sum_{r_g} y_{k,t-1}} \quad (2)$$

och  $y_{k,t-1}$  är det inkomna industrivärdet för företag  $k$  år  $t-1$ .

$r_g$  är svarsmängden av de gemensamma företag som svarat vid år  $t$  och  $t-1$  för branschgrupp  $g$ .

$$r_g = r_{g,t} \cap r_{g,t-1}$$

För att beräkna kvoten  $\hat{R}_g$  ska minst 1 företag tillhöra svars-  
mängden  $r_g$  och ha ett individuellt utvecklingstal som ligger inom  
intervallet 0,6-1,4.

Kvoten  $\hat{R}_g$  beräknas för branschgrupper indelade enligt 5-siffer  
SNI, 4-siffer SNI, 3-siffer SNI och 2-siffer SNI. Finns inget utveck-  
lingstal för branschgrupp  $g$  enligt 2-siffer SNI sätts  $\hat{R}_g = 1$ .

För ett företag  $k$  som tillhör branschgrupp  $g$  enligt 5-siffer SNI som  
ska imputeras så hämtas kvoten  $\hat{R}_g$  från 5-siffer SNI och  $\hat{y}_{k,t}$  beräk-  
nas enligt uttryck 1. Finns ingen kvot enligt 5-siffer SNI så tas  
kvoten enligt 4-siffer SNI osv.

För information om SNI (Svensk Näringsgrensindelning) se  
<http://www.foretagsregistret.scb.se/sni>

## Skattning av industrivärde utifrån SRU- och FEK- uppgifter (4)

Standardiserade RäkenskapsUtdrag (SRU) innehåller årsuppgifter  
från företagens skattedeclaration. Huvudsakligen består utdraget av  
företagens resultat- och balansräkning. Utdraget levereras till SCB  
från Skatteverket. Merparten av uppgifterna kommer till SCB i slutet  
på augusti året efter referensåret. En mindre leverans tillkommer i  
slutet på december.

Med nettoomsättningen från SRU som hjälpvariabel skattas sedan  
ett företags totala industrivärde med kvotestimationsteknik. Kvoten  
beräknas som andelen industri i förhållande till nettoomsättningen,  
d v s en proportion,  $p$ . Underlaget till att skatta  $p$ , hämtas från de  
blankettundersökta företagen i undersökningen Företagens Ekonomi  
(FEK).

### Skattning utifrån SRU- och FEK-uppgifter år $t$ (4b)

FEK:ens industripopulation  $U$  är indelade i  $H$  stycken strata,  
 $U_1, \dots, U_h, \dots, U_H$ . Stratumindelningen utgår från

Nationalräkenskaperens årsbranscher. Ett stratum  $U_h$  innehåller  $N_h$   
stycken företag från vilka  $n_h$  stycken dras enligt urvalsdesignen  
nedan.

$\pi_k$  som är första ordningens inklusionssannolikhet för att företag  $k$  ska ingå i FEK urvalet, bestäms enligt

$$\pi_k = \frac{n_h z_k}{\sum_{U_h} z_k}$$

om företag  $k$  tillhör stratum  $h$ .

$\pi_k$  bestäms proportionellt mot variabeln  $z_k$  som är totala intäkter och totala kostnader för företag  $k$  enligt SRU.

Modellgruppsindelningen utgår från stratumindelningen uppdelade på två storleksgrupper. Den första gruppen innehåller företag som har  $\pi_k = 1$  medan den andra gruppen innehåller företag som har  $\pi_k < 1$ .

För ett företag  $k$  skattas det totala industrivärdet,  $y_k$ , i IVP enligt

$$\begin{cases} \hat{y}_k = \hat{p}_g \times x_k \text{ om företag } k \text{ tillhör modellgrupp 2 och storleksgrupp 2} \\ \hat{y}_k = y_k^* \text{ om företag } k \text{ tillhör storleksgrupp 1} \end{cases} \quad (3)$$

$$\text{Där } \hat{p}_g = \frac{\sum_{r_g} \frac{y_k^*}{\pi_k}}{\sum_{r_g} \frac{x_k}{\pi_k}} \quad (4)$$

och  $x_k$  är nettoomsättning för företag  $k$  enligt SRU och  $y_k^*$  är industrivärdet för företag  $k$  enligt FEK.

Ett problem är dock att vid septemberpubliceringen saknas ett tillräckligt underlag av inkomna blanketter i FEK för att beräkna  $\hat{p}_g$ .

$p_g$  skattas då utifrån fjolårets industriandel, d v s  $\hat{p}_{g,t} = \hat{p}_{g,t-1}$ .

### Skattning utifrån SRU- och FEK-uppgifter år t-1 (4a)

I juni finns varken tillgång till årets SRU-uppgifter ( $x_k$ ) eller ett tillräckligt underlag av inkomna blanketter i FEK för att beräkna  $\hat{p}_g$ .

Tillvägagångssättet för att skatta  $y_k$  är att utgå från fjolårets skattade industrivärde,  $\hat{y}_{k,t-1}$  som beräknas analogt med uttryck 3.

Sedan framskrivs  $\hat{y}_{k,t-1}$  med utvecklingstalet från uttryck 2 vilket ger skattningen för år  $t$  enligt

$$\hat{y}_{k,t} = \hat{R}_g \times \hat{y}_{k,t-1} \text{ om företag } k \text{ tillhör branschgrupp } g.$$

## Omsättningsuppgifter i momsdeklarationen från Skatteverket (5)

Varje månad redovisar alla företag som har en årlig omsättning som överstiger 1 mnkr in- och utgående moms i en momsdeklaration till Skatteverket. De större företagen (omsättning större än 40 mnkr) ska redovisa momsuppgifter senast 25 dagar efter referensmånadens utgång. Övriga företag ska redovisa momsuppgifter senast 40 dagar efter referensmånadens utgång. Ett företags omsättning beräknas utifrån vad företaget har redovisat i utgående moms och export enligt deklarationen.

Låt  $m_k$  beteckna nettoomsättning från momsdeklarationen för företag  $k$ . Det totala industrivärdet i IVP skattas då enligt

$$\hat{y}_k = m_k$$

Det innebär ett antagande om att all nettoomsättning för företag  $k$  utgörs av industriverksamhet. Uppgifter från momsdeklarationen är den sista datakällan som används vilket innebär att det mest är små företag som får denna källa som underlag. För dessa företag är det ett rimligt antagande att all verksamhet består av industri.

# Bilaga 2

## Tabellsammanställning

**Tabell 2**

**Korstabell över delpopulationer för publiceringstillfälle 1 och 2, år 2004. Värden i cellerna är i tusentals kronor (tkr)**

P1 (juni)		P2 (september)				
		Delpop p2				
Delpop p1	Skattad Parameter	1 (svarande)	2 (bortfall)	3 (varannan års)	4 (cutoff)	Totalt
1 (svarande)	<b>Differens</b>	<b>92 210</b>	-112 271			<b>-20 061</b>
	Total P1	1 139 665 943	1 030 261			1 140 696 204
	Antal företag	3 464	3			3 467
2 (bortfall)	<b>Differens</b>	<b>-1 119 238</b>	<b>-657 223</b>			<b>-1 776 461</b>
	Total P1	9 275 073	61 944 136			71 219 209
	Antal företag	41	555			596
3 (varannan års)	<b>Differens</b>			<b>40 590</b>		<b>40 590</b>
	Total P1			40 506 789		40 506 789
	Antal företag			774		774
4 (cutoff)	<b>Differens</b>				<b>-6 804 690</b>	<b>-6 804 690</b>
	Total P1				141 340 214	141 340 214
	Antal företag				57 273	57 273
Undertäckning	<b>Differens</b>	18 814				18 814
	Total P1	0				0
	Antal Företag					
Totalt	<b>Differens</b>	<b>-1 008 214</b>	<b>-769 494</b>	<b>40 590</b>	<b>-6 804 690</b>	<b>-8 541 808</b>
	Total P1	1 148 941 015	62 974 397	40 506 789	141 340 214	1 393 762 416
	Antal företag	3 505	558	774	57 273	62 110



**Tabell 2**  
**Korstabell över delpopulationer för publiceringstillfälle 1 och 2, år 2005. Värderna i cellerna är i tusentals kronor (tkr)**

P1 (juni)		P2 (september)				
		Delpop p2				
Delpop p1	Skattad Parameter	1 (svarande)	2 (bortfall)	3 (varannan års)	4 (cutoff)	Totalt
1 (svarande)	<b>Differens</b>	<b>3 016 895</b>				<b>3 016 895</b>
	Total P1	1 168 769 224				1 168 769 224
	Antal företag	3 070				3 070
2 (bortfall)	<b>Differens</b>	<b>-352 819</b>	<b>593 672</b>			<b>240 853</b>
	Total P1	33 045 930	23 054 765			56 100 695
	Antal företag	180	392			572
3 (varannan års)	<b>Differens</b>			<b>46 196</b>		<b>46 196</b>
	Total P1			62 124 282		62 124 282
	Antal företag			1 096		1 096
4 (cutoff)	<b>Differens</b>				<b>-5 439 421</b>	<b>-5 439 421</b>
	Total P1				143 031 220	143 031 220
	Antal företag				58 734	58 734
<b>Totalt</b>	<b>Differens</b>	<b>2 664 076</b>	<b>593 672</b>	<b>46 196</b>	<b>-5 439 421</b>	<b>-2 135 477</b>
	Total P1	1 201 815 154	23 054 765	62 124 282	143 031 220	1 430 025 420
	Antal företag	3 250	392	1 096	58 734	63 472

Tabell 4

**Korstabell över källorna för företag under cutoff-gränsen för publiceringstillfälle 1 och 2, år 2004. Värdena i cellerna är i tkr**

P1 (juni)		P2 (sept)				
		Källa p2				
Källa p1	Skattad Parameter	Manuell imputering	Sru (t)	Moms	Källa saknas	Totalt
Manuell imputering	<b>Differens</b>	<b>0</b>				<b>0</b>
	Total P1	64 083				64 083
	Antal företag	37				37
Sru (t-1)	<b>Differens</b>		<b>777 970</b>	<b>-3 981 001</b>	<b>-4 961 702</b>	<b>-8 164 732</b>
	Total P1		121 174 948	6 070 106	4 961 702	132 206 755
	Antal företag		43 840	786	2 905	47 531
Moms	<b>Differens</b>		<b>-242 435</b>	<b>0</b>		<b>-242 435</b>
	Total P1		3 574 332	5 495 044		9 069 376
	Antal företag		1 860	1 266		3 126
Källa saknas	<b>Differens</b>		<b>1 117 607</b>		<b>0</b>	<b>1 117 607</b>
	Total P1		0		0	<b>0</b>
	Antal företag		2 216		4 363	6 579
Totalt	<b>Differens</b>	<b>0</b>	<b>2 138 012</b>	<b>-3 981 001</b>	<b>-4 961 702</b>	<b>-6 804 690</b>
	Total P1	64 083	124 749 280	11 565 150	4 961 702	141 340 214
	Antal företag	37	47 916	2 052	7 268	57 275

**Tabell 4**  
**Korstabell över källorna för företag under cutoff-gränsen för publiceringstillfälle 1 och 2, år 2005. Värdena i cellerna är i tkr**

P1 (juni)		P2 (sept)				
		Källa p2				
Källa p1	Skattad Parameter	Manuell imputering	Sru (t)	Moms	Källa saknas	Totalt
Manuell imputering	<b>Differens</b>	<b>0</b>				<b>0</b>
	Total P1	1 364 608				1 364 608
	Antal företag	65				65
Sru (t-1)	<b>Differens</b>	<b>-160 343</b>	<b>869 049</b>	<b>-2 633 729</b>	<b>-3 417 481</b>	<b>-5 342 503</b>
	Total P1	160 343	124 659 204	4 083 284	3 417 481	132 320 312
	Antal företag	4	44 490	745	3 112	48 351
Moms	<b>Differens</b>	<b>-199 633</b>	<b>-1 054 773</b>	<b>0</b>		<b>-1 254 406</b>
	Total P1	199 633	8 255 503	891 164		9 346 300
	Antal företag	3	2 589	364		2 956
Källa saknas	<b>Differens</b>		<b>1 157 488</b>		<b>0</b>	<b>1 157 488</b>
	Total P1		0		0	0
	Antal företag		3 575		3 787	7 362
Totalt	<b>Differens</b>	<b>-359 976</b>	<b>971 764</b>	<b>-2 633 729</b>	<b>-3 417 481</b>	<b>-5 439 421</b>
	Total P1	1 724 584	132 914 707	4 974 448	3 417 481	143 031 220
	Antal företag	72	50 654	1 109	6 899	58 734

**Tabell 5**  
**Korstabell över källorna för företag under cutoff-gränsen för publiceringstillfälle 1 och 2 för år 2004. Värdena i cellerna är i tkr**

P1 (juni)		P2 (sept)					Totalt
		Källa p2					
Källa p1	Skattad Parameter	Manuell imputering	Sru (t)	Moms	Källa saknas		
Manuell imputering	<b>Differens</b>	<b>0</b>				<b>0</b>	
	<b>Korrigerad differens</b>	<b>0</b>				<b>0</b>	
	Antal företag	37				37	
Sru (t-1)	<b>Differens</b>		<b>777 970</b>	<b>-3 981 001</b>	<b>-4 961 702</b>	<b>-8 164 732</b>	
	<b>Korrigerad differens</b>		<b>1 116 620</b>	<b>-779 651</b>	<b>-1 620 026</b>	<b>-1 283 056</b>	
	Antal företag		43 840	786	2 905	47 531	
Moms	<b>Differens</b>		<b>242 435</b>	<b>0</b>		<b>-242 435</b>	
	<b>Korrigerad differens</b>		<b>242 435</b>	<b>0</b>		<b>-242 435</b>	
	Antal företag		1 860	1 266		3 126	
Källa saknas	<b>Differens</b>		<b>1 117 607</b>		<b>0</b>	<b>1 117 607</b>	
	<b>Korrigerad differens</b>		<b>1 117 607</b>		<b>0</b>	<b>1 117 607</b>	
	Antal företag		2 216		4 363	6 579	
Totalt	<b>Differens</b>	<b>0</b>	<b>2 138 012</b>	<b>-3 981 001</b>	<b>-4 961 702</b>	<b>-6 804 690</b>	
	<b>Korrigerad differens</b>	<b>0</b>	<b>2 476 662</b>	<b>-779 651</b>	<b>-1 620 026</b>	<b>-76 986</b>	
	Antal företag	37	47 916	2 052	7 268	57 275	

**Tabell 5**  
**Korstabell över källorna för företag under cutoff-gränsen för publiceringstillfälle 1 och 2 för år 2005. Värdena i cellerna är i tkr**

P1 (juni)		P2 (sept)					Totalt
		Källa p2					
Källa p1	Skattad Parameter	Manuell imputering	Sru (t)	Moms	Källa saknas		
Manuell imputering	<b>Differens</b>	<b>0</b>					<b>0</b>
	<b>Korrigerad differens</b>	<b>0</b>					<b>0</b>
	Antal företag	65					65
Sru (t-1)	<b>Differens</b>	<b>-160 343</b>	<b>869 049</b>	<b>-2 633 729</b>	<b>-3 417 481</b>		<b>-5 342 503</b>
	<b>Korrigerad differens</b>	<b>-160 343</b>	<b>1 238 055</b>	<b>-466 857</b>	<b>-713 468</b>		<b>-102 613</b>
	Antal företag	4	44 490	745	3 112		48 351
Moms	<b>Differens</b>	<b>-199 633</b>	<b>-1 054 773</b>	<b>0</b>			<b>-1 254 406</b>
	<b>Korrigerad differens</b>	<b>-199 633</b>	<b>-1 054 773</b>	<b>0</b>			<b>-1 254 406</b>
	Antal företag	3	2 589	364			2 956
Källa saknas	<b>Differens</b>		<b>1 157 488</b>		<b>0</b>		<b>1 157 488</b>
	<b>Korrigerad differens</b>		<b>1 157 488</b>		<b>0</b>		<b>1 157 488</b>
	Antal företag		3 575		3 787		7 362
Totalt	<b>Differens</b>	<b>-359 976</b>	<b>971 764</b>	<b>-2 633 729</b>	<b>-3 417 481</b>		<b>-5 439 421</b>
	<b>Korrigerad differens</b>	<b>-359 976</b>	<b>1 340 770</b>	<b>-466 857</b>	<b>-713 468</b>		<b>-199 531</b>
	Antal företag	72	50 654	1 109	6 899		58 734



# In English

## Summary

The five main components in terms of quality of official statistics are contents, timeliness, accuracy, comparability and availability. This study focus on timeliness in one of Statistics Sweden's surveys, Industrial Production of goods.

The survey publishes national statistics nine months after the end of the reference time. In this study we investigate the effects of the estimates of the total industrial production on a three months earlier publishing. We categorize some of the survey's most important sources of error – namely editing, nonresponse and imputation . The effects from each sources of error are estimated.

The main results from the study are that an earlier publishing has effects on the estimates and it seems to be the imputation effect that is of considerable size. The methods to impute values seems to be sensitive to which auxiliary information the survey has access to by the time of publishing. A critical group of enterprises seems to be more vulnerable to the choice of imputation method. An alternative imputation method was tested on this group of enterprises. This approach uses more current data as auxiliary information. The alternative approach reduces the imputation effect for this group of enterprises and gives more accordant estimates between earlier and later time of publishing.

## Disclaimer

The series Research and Development – Methodology reports from Statistics Sweden is published by Statistics Sweden and includes results on development work concerning methods and techniques for statistics production. Contents and conclusions in these reports are those of the author(s).

## Kartläggning av felkällor för bättre aktualitet

De fem huvudkomponenterna i kvalitetsbegreppen för officiell statistik är innehåll, aktualitet, tillförlitlighet, jämförbarhet och tillgänglighet. Denna rapport sätter fokus på begreppet aktualitet i en av SCB:s undersökningar, Industrins Varuproduktion.

I rapporten studeras effekterna på skattningarna av en tidigare-läggning av publiceringen på tre månader. I studien görs en uppdelning på några av undersökningens viktigaste felkällor – nämligen granskning, bortfall och imputering.

ISSN 1653-7149

### **Publikationstjänsten:**

E-post: [publ@scb.se](mailto:publ@scb.se), tfn: 019-17 68 00, fax: 019-17 64 44. Postadress: 701 89 Örebro.

**Information och bibliotek:** E-post: [information@scb.se](mailto:information@scb.se), tfn: 08-506 948 01, fax: 08-506 948 99.  
Försäljning över disk, besöksadress: Biblioteket, Karlavägen 100, Stockholm.

### **Publication services:**

E-mail: [publ@scb.se](mailto:publ@scb.se), phone: +46 19 17 68 00, fax: +46 19 17 64 44. Address: SE-701 89 Örebro.

**Information and Library:** E-mail: [information@scb.se](mailto:information@scb.se), phone: +46 8 506 948 01, fax: +46 8 506 948 99.  
Over-the-counter sales: Statistics Sweden, Library, Karlavägen 100, Stockholm, Sweden.