

# Brus i FoU – ny metod att skydda data



Sveriges officiella  
statistik

# Brus i FoU – ny metod att skydda data

Producent SCB, Statistiska centralbyrån  
Avdelningen för ekonomisk  
statistik och analys  
171 54 Solna  
010-479 40 00

Förfrågningar Nils Adriansson  
010-479 42 78  
nils.adriansson@scb.se

Radka Sabolová  
010-479 61 37  
radka.sabolova@scb.se

Du får kopiera och på annat sätt mångfaldiga innehållet.  
Vi vill dock att du uppger källa på följande sätt:  
Källa: SCB, Brus i FoU – ny metod att skydda data.

URN:NBN:SE:SCB-2024-UFFTBR2024\_pdf

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

## Bakgrund

Uppgifterna i tabellerna för undersökningarna Forskning och utveckling (FoU) i företagssektorn samt privata icke-vinstdrivande sektorn skyddas genom undertryckning, så kallad prickning, av celler som innehåller potentiellt känsliga data. En konsekvens av detta är att viss statistisk information inte kan tillhandahållas och användares behov kan därmed inte alltid tillgodoses i den officiella statistiken. Särskilt tabeller som redovisar statistik efter näringsgren, eller näringsgren i kombination med andra redovisningsgrupper som storleksklass, undertrycks vilket minskar statistikens användbarhet och relevans.

För att undvika prickning av celler inhämtas medgivanden från viktiga uppgiftslämnare för publicering av celler innehållandes deras uppgifter. Det innebär att fler kontakter behöver tas med organisationen som då behöver ta ställning till om de lämnade uppgifterna kan leda till men eller skada för organisationen. Dessa medgivanden lämnas ofta obesvarade eller nekar publicering av celler med potential att röja organisationens uppgifter.

Ett omfattande arbete med röjandekontroll görs för att säkerställa att inga organisationer kan röjas. Primärundertryckning sker genom den så kallade p%-regeln och leder även till sekundärundertryckning för att undvika att delar röjs då totaler redovisas. För företagssektorn görs även kontroller av data som redovisas i FoU i företagssektorn, Innovation i företagssektorn och hos Eurostat där samma variabler publiceras under olika aggregeringar, vilket leder till ytterligare undertryckning.

Mot bakgrund av detta är FoU i företagssektorn samt FoU i privata icke-vinstdrivande sektorn de första undersökningarna på SCB där en perturbativ metod istället för undertryckning används på magnitudtabeller.

## Beskrivning av metoden

EZS-metoden är en enkel perturbativ metod för magnitudtabeller som utvecklades av Evans, Zayatz och Slanta (1998) på US Census Bureau. Eftersom metoden adderar ett ”brus” på mikronivå (dvs. vi ändrar insamlade uppgifter) behålls additivitet i hierarkier och kopplingar mellan tabellerna.

Varje objekt (organisation) får en positiv eller negativ riktning, -1 eller +1, och sedan en slumpmässig storlek på bruset som simuleras utifrån en vald fördelning (i procentenheter). Dessa parametrar förblir konfidentiella. Brusade värden beräknas enligt formeln

$$\text{brusade värde} = \text{värde} * (1 + \text{riktning} * \text{brus}/100),$$

där samma storlek och riktning på bruset används för alla uppgifter som objektet lämnat, dvs. för alla variabler och i alla tabeller. Fördelningen

av riktningar är symmetrisk runt 0 så att den inte introducerar skevheter.

Celler som innehåller bara en observation eller celler med ett dominant bidrag förväntas få större brus eftersom vi strävar efter att skydda individuella objekt i dessa celler. I celler med många små bidrag blir effekten av brus mindre och cellens totala brusade värde förväntas hamna närmare det obrusade värdet.

För att minska den totala mängden brus som lagts till data tillämpades en balanseringsprocedur (föreslogs i Massel och Funk (2007)). Balansering tillämpas endast på celler som är icke-känsliga enligt den använda röjandekontrollregeln och inte försämrar skydds nivån.

Balansering kräver val av en tabell och en variabel som styr balanseringen. Balansering innebär att man minimerar bruset i celler genom att ändra riktningen på bruset för vissa av de objekt som ingår i cellen. Den nya riktningen tillämpas sedan på alla uppgifter som lämnats av objektet. Proceduren tillämpas på den mest detaljerade redovisningsnivån.

Balanseringsproceduren använder sig av 'girig algoritm' för att reducera brus i icke-känsliga celler. Bidragen i respektive cell sorteras från största till minsta. Det största bidraget behåller sin slumpmässigt tilldelade riktning. Resterande objekts riktningar kan ändras så att det totala bruset vid varje steg minimeras och den kumulativa summan i cellen blir närmare den ursprungliga obrusade kumulativa summan. Detta uppnås genom att välja en riktning på dessa objekt som ska vara den motsatta riktningen än riktningen på det kumulativa bruset i cellen. I osäkra celler (primära undertryckningar) används den obalanserade versionen av metoden. Celler som skulle sekundärundertryckas är inte känsliga och därför balanseras dessa. Observera att balanseringsproceduren utförs i en av tabellerna, men dess resultat påverkar alla tabeller.

## Exempel

Detta korta exempel illustrerar balanseringsproceduren och EZS-metoden. Vi tittar närmare på en cell med bidrag från 5 olika objekt. Alla objekt tilldelades slumpmässiga riktningar och storlek på bruset. Obrusade cellens total är lika med 2 000, och brusade obalanserade totalen är lika med 2 068,08, en ökning med 3,4 procent. Enligt röjandekontrollregeln är den här cellen icke-känslig och därför är det önskvärt att ändra riktningar så att skattningen med mindre brus skulle kunna publiceras. De två största bidragen behåller sina motsatta riktningar eftersom dessa leder till en mindre brusad kumulativ summa i cellen. Däremot ändras riktningen för tredje objektet F3 så den brusade summan blir närmare den obrusade, ursprungliga, summan. De två sista objekten, F4 och F5, som bidrar till cellen behåller sina

ursprungliga slumpmässiga riktningar. Den balanserade totalen blir 1 990,92 som motsvarar en minskning med 0,45 procent relativt brus.

**Tabell 1. Tillämpning av EZS metod och balanseringsproceduren, observera att objekt F3 bytt riktning.**

Objekt	Brus	Riktning (ursprunglig)	Värde	Kumulativ summa	Obalanserad	Riktning +1	Riktning -1	Ny riktning
F1	10,94	1	1 000	1 000	1 109,40			
F2	13,77	-1	450	1 450	1 497,44	1 621,37	1 497,44	-1
F3	12,86	1	300	1 750	1 836,02	1 836,02	1 758,86	-1
F4	11,63	-1	200	1 950	2 012,76	1 982,12	1 935,60	-1
F5	10,65	1	50	2 000	2 068,08	1 990,92	1 980,27	1

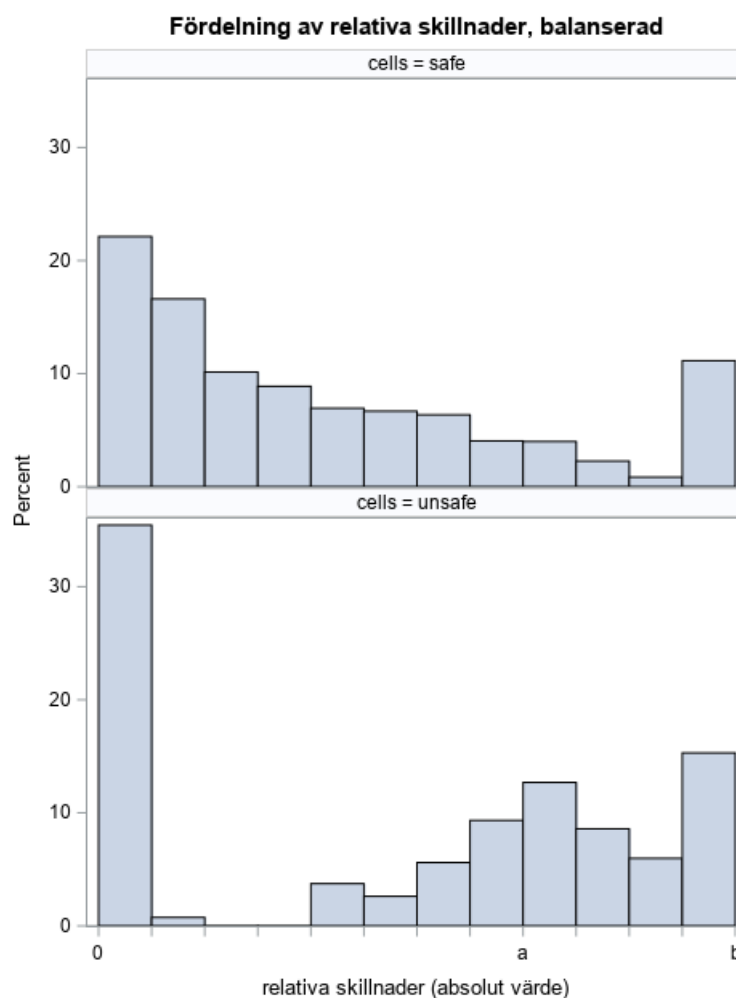
## Resultat och tillämpning

EZS-metoden implementerades och testades på 2021 data från FoU undersökningen i företagssektorn. Brusets storlek simulerades från vald fördelning med värden mellan *a* och *b*. I beredningsarbetet har olika styrande variabler testats, dels de som hade flest osäkra celler, dels de som representerar betydande del av publicerade totala utgifter för egen FoU. Olika val av variabler ger olika resultat beroende på värden och mängden säkra och osäkra celler, vilket visar att resultatet av balanseringsproceduren är mycket känsligt för dessa faktorer.

Som utgångspunkt undersökte SCB EZS-metodens effekt på en av undersökningens huvudvariabler, totala utgifter för egen FoU, och fann att skattningarna skilde sig åt betydligt beroende på val av metod (balanserad eller obalanserad), variabel för balansering och graden av avrundning. För att utvärdera metoder har SCB tittat på fördelningen av absoluta värden av relativa skillnader mellan de ursprungliga skattningarna och skattningarna baserade på brusade värden. Detta gjordes separat för säkra respektive osäkra celler. För osäkra celler vill vi ha större skillnader för att de ska ha ett tillräckligt skydd och för säkra celler vill vi ha mindre skillnader för att kunna publicera så nära deras observerade värden som möjligt. Effekten av brus varierades, vid obalanserad metod har de totala utgifterna blivit 4,8 procent högre, olika val av balanseringsvariabler resulterades in i ökning mellan 0,7 and 5,6 procent.

Följande histogram visar fördelningen av absoluta värden av relativa skillnader för ett val av balanseringsvariabel. X-axeln representerar procentandelen brus som läggs till celler (i absoluta värden), y-axeln representerar andelen celler. I den övre panelen finns histogram för säkra celler (inkl. celler som skulle sekundärundertryckas) som visar att dessa celler fick mindre brus. Nästan en fjärdedel av celler fick nästan inget brus. Ungefär 15 procent av säkra celler fick större mängd brus. Histogrammet för osäkra celler (nedre diagrammet) har en annan fördelning: stort antal celler fick åtminstone *a* procent brus, dvs.

minimum brus som tillämpas på varje observation. Effekten av bruset minskar för en stor andel osäkra celler vid avrundning till miljoner kronor, det gäller framför allt celler med små värden. Deras brusade avrundade värden skilde sig inte från deras avrundade obrusade värden (se den högsta stapeln vid  $x = 0$ ).



Figur 1. Fördelningen av relativa skillnader, balanserad version av EZS-metoden.

## Slutsats

Metoden testades på data från FoU undersökning i företagssektorn 2021 och tillämpades på 2023 data för företagssektorn samt privat icke-vinstdrivande sektorn som publicerades hösten 2024. Metoden möjliggör spridning av tabeller med brusade värden utan att undertrycka några celler på ett konsekvent sätt samt är enkel att implementera utan behov av någon specialiserad programvara. Eftersom brus läggs på underliggande mikrodata fungerar metoden bra även för tabeller med hög dimension eller tabeller med hierarkier och är

mycket mindre tidskrävande än vanlig röjandekontroll och cellundertryckning.

Resultatet från implementering på 2021 års data har inget informativt värde om data som publicerades för referensår 2023. Tilldelade brus kan skilja sig och relativa skillnader beror på observerade värden och andra faktorer. Mellan 2021 och 2023 har det också skett andra förändringar i undersökningens design som påverkar det publicerade resultatet.

# Referenser

T. Evans, L. Zayatz and J. Slanta, Using Noise for Disclosure Limitation of Establishment Tabular Data, *Journal of Official Statistics*, Vol 14, No. 4, 1998, pp. 537-551.

P. B. Massell and J. M. Funk, Protecting the Confidentiality of Tables by Adding Noise to the Underlying Microdata, paper presented at the ICES-III, June 18-21, 2007, Montreal, Quebec, Canada.



