

Tema: Nye Datamuligheder

Kan vi gøre det billigere og bedre?

En undersøgelse af mulighederne for at anvende administrative data i beregningen af det danske industriproduktionsindeks.

Søren Kristensen,
Chefkonsulent
Konjunkturstatistik
skr@dst.dk
Danmarks Statistik

Abstract

De nationale statistikbureauer står over for store udfordringer de kommende år. Samtidig med, at der er et stigende behov for mere, nyere og hurtigere statistik, er der fokus på reduktion af respondentbyrden og effektivisering af dataindsamlingen. En måde at forene disse krav på er at gøre større brug af administrative data. I Danmark gør vi os overvejelser om brug af momsdata (VAT) i forbindelse med beregning af industriproduktionsindekset (IP), der i dag alene beregnes på baggrund af stikprøvedata. Udfordringerne i forbindelse med anvendelse af administrativ data er velkendte. Grundlæggende hænger de sammen med, at dataindsamlingen ikke er i statistikbureauets egne hænder og ikke har statistik som formål. De største udfordringer i forhold til IP er, at der er forskel på de administrative og de statistiske enheder, samt at der mangler information om variabelen 'ændring af lager'. En metode til at imødegå dette er at kombinere administrativ data med stikprøvedata.

1. Indledning

Krav om hurtigere og mere effektiv statistikproduktion

Der er ingen tvivl om at der i disse år er et stigende behov for mere, hurtigere og bedre statistik. Dette blev særligt tydeligt i forbindelse med finanskrisen, hvor man næsten ikke kunne få konjunkturindikatorerne hurtigt nok. Samtidig er de nationale statistikbureauer generelt underlagt effektiviseringskrav, som næppe bliver mindre i takt med at der gennemføres statslige spareplaner i kølvandet på finanskrisen. Der har også i længere tid været fokus på at reducere den byrde virksomhederne pålægges, når vi indsamler informationer til statistik. Kort fortalt skal vi producere bedre statistik, vi skal producere det billigere og helst uden at genere erhvervslivet.

Administrativ data – en mulig løsning

En større brug af administrative data kan være en måde at få disse mål til at hænge sammen. Her er vi så også godt hjulpet af, at den teknologiske udvikling har medført, at der indsamles mere information digitalt og der dermed er mere administrativ data til rådighed.

Industriproduktionsindekset (IP) er ofte blevet nævnt som et eksempel på en statistik, hvor der burde være mulighed for at anvende administrative data, nærmere betegnet momsdata (VAT) i stedet for at indsamle data fra virksomhederne. Selvom det virker indlysende at bruge momsdata, er det ikke så enkelt, fordi de data vi kan hente via moms ikke helt svarer til det, vi har brug for. I det har paper beskrives de foreløbige resultater af en undersøgelse af i hvor høj grad og hvordan vi kan bruge momsdata i stedet for spørgeskema data til IP.

2. Om industriproduktionsindekset

Industriproduktionsindekset (IP) er en af de væsentlige konjunkturindikatorer, og har som formål er at give et billede af udviklingen i industrien fra måned til måned. Indekset adskiller sig fra et omsætningsindeks ved at fokus ikke er på, hvad der bliver solgt, men derimod på hvad der bliver produceret. Noget af det der bliver produceret kommer direkte på lager og noget af det der solgt hentes fra lageret.

Den operationelle definition af IP Ifølge den teoretiske definition er det egentlig værditilvæksten, der skulle måles med IP. Det er dog de færreste lande, hvis nogle overhovedet, der i praksis måler værditilvækst. Nogle lande indsamler månedlige data om produktion af industrivarer, det der svarer til Prodcum, og bruger det til industriproduktionsindekset. Andre, heriblandt Danmark, beregner produktionen som en funktion af den månedlige omsætning og ændringer i lagerværdien. Både lagerværdi og den månedlige omsætning deflateres.

3. Udfordringer ved brug af momsdata

Forskel i definition af omsætning Vi kan imidlertid ikke uden videre anvende moms i stedet for de indsamlede omsætningsdata. Mens momsdata med nogle enkelte undtagelser omfatter al den omsætning en virksomhed har, så er det et noget snævrere omsætningsbegreb, der anvendes til beregning af industriproduktionsindekset. En af de store forskelle mellem omsætningsbegreberne er f.eks. at omsætning fra videresalg af varer, dvs. varer som ikke er produceret eller forædlet af virksomheden selv, af gode grunde ikke skal med i et produktionsindeks. I nogle brancher udgør dette en relativ stor andel af den samlede omsætning.

Mangler information om lager En anden stor udfordring er, at ændringer i lagerværdi, der er en central variabel til beregning af IP på ingen måde kan indhentes via momsdata.

Forskel mellem indsamlingsenheder En tredje udfordring er, at de enheder der indsamles data for er forskellige. Til IP indsamler vi data fra de såkaldte faglige enheder (KAU), som er rent statistiske enheder, der er dannet med det formål at lave en mere præcis kobling mellem data og branche. Momsdata er derimod koblet til juridiske enheder og deres brancheplacering. For mange virksomheder, særligt de mindre virksomheder, er der reelt ikke forskel mellem den juridiske enhed og den faglige enhed. Men for de større og mere komplekse virksomheder, som typisk fylder meget i statistikkerne, vil det som regel være tilfældet, at de består af et antal forskellige faglige enheder, hvoraf nogle endog kan ligge uden for industrien.

Opsummering – hvilke udfordringer er der? For at kunne anvende momsdata til IP er det derfor nødvendigt at overveje:

- Om vi kan justere momstallene så definitionen af omsætningen svarer til det der skal bruges til IP.
- Om vi kan indsamle information om lagerændringer fra andre kilder
- Om man kan tage højde for at der er forskelle mellem de enheder, der indberettes data for.

4. Hvilke muligheder er der for tilpasning af data?

Justering af momsdata For at starte med det andet punkt, så har vi i prodcom statistikken (der indsamles kvartalsvis i Danmark) en kilde til at foretage en justering af momsdata. I prodcom indsamles oplysninger om såvel salg af egne varer som omsætning ved videresalg af handelsvarer. Da vi bruger prodcom som stikprøveramme og til opregning, er der en god sammenhæng mellem IPs enheder og prodcoms enheder. Derfor kan vi for hver enhed beregne en faktor som angiver, hvor stor en andel salg af egne varer udgør i forhold til den samlede omsætning.

Datakvalitet Den omregningsfaktor man kan beregne er naturligvis usikker fordi den ikke alene skal dække alle tre måneder i et kvartal, men også skal dække en række måneder forud.

Herudover viste det sig ved en nærmere analyse af data, at der på enhedsniveau godt kan være meget stor forskel mellem momsindberetningerne og de tal der angives for omsætningen i forbindelse med dataindsamlingen til IP. Der kan være flere forklaringer på dette, f.eks. forskellig periodisering eller uklarhed fra indberetternes

side om definitionen af omsætning. Vi har ikke undersøgt nærmere hvilke tal, der er mest retvisende, eller hvad der konkret er årsagen til afvigelserne, men det er klart, at det kunne være nyttigt at komme mere til bunds i den problemstilling.

God mulighed for estimation af omsætningsvariablen

Selv med de begrænsninger der nu er, er der dog næppe tvivl om at det ville være muligt at finde en god løsning for selve omsætningsvariablen. Men de to andre udfordringer er umiddelbart lidt sværere at løse.

Intet brugbart alternativ kilde til lagerændringer

Der er pt. ikke en god alternativ kilde til månedlige lagerændringer. Ganske vist udarbejder vi en spørgeskemabaseret lagerværdistatistik, men det er en kvartalstatistik, så den kan i bedste fald kun bruges til at benchmarke IPs egne månedlige lagerværdiindberetninger mod.

Omregning fra juridiske til faglige enheder

Der er heller ikke nogen rigtig god metode til at fordele momsomsætningen som vi kun har på virksomhedsniveau på de faglige enheder. Vi kan finde ud af, hvor mange faglige enheder, der er tilknyttet de enkelte virksomheder, men antallet af ansatte er umiddelbart den bedste variabel vi kan anvende til fordeling af omsætning og det giver selvsagt ikke et særligt pålideligt resultat.

To muligheder for anvendelse af momsdata

Hvis vi ønsker at anvende momsdata til beregningen af IP, er der således generelt set to veje at gå.

Vi kan enten bruge momsdata suppleret med indsamlede oplysninger om lagerændringer og fordeling mellem de faglige enheder, eller vi kan vælge at skræddersy stikprøven, så vi kun indsamler data fra de virksomheder, vi har de største problemer med at anvende momsdata for. Da det næppe bliver anset som en reducere af respondentbyrden, hvis der blot fjernes et par felter fra spørgeskemaet, så er det den sidstnævnte mulighed vi har valgt.

Fokusering på mindre virksomheder med en faglig enhed

De virksomheder vi har udvalgt som dem hvor vi bedst kan anvende momsdata er mindre virksomheder, der kun består af en faglig enhed. Dermed undgår man problematikken om fordelingen af momsomsætningen, og da disse virksomheder vægter relativt lidt i det samlede indeks er fylder afvigelse på individniveau mindre.

5. Analyse af data

Nøgletal for stikprøven

I tabellen nedenfor vises nogle nøgletal for vores opregnede stikprøve fra 2009. Vi har delt virksomhederne op efter årsomsætning og om de består af en eller flere faglige enheder. Tabellen viser antallet af virksomheder i de fem grupper, samt hvor stor en del af IP omsætningen og forskellen mellem moms og IP omsætningen de tegner sig for. Tabellen viser også andelen af de samlede lagerændringer over perioden. Da summerede lagerændringer har det med at gå mod nul over tid, er summen af lagerændringer her beregnet ud fra den numeriske værdi alene.

Nøgletal for IP 2009 (i pct.) – fordelt efter antal faglige enheder og omsætning

Tabel 1

Virksomhedstype	Årlig omsætning mio DKK	Antal	2009		
			IP omsætning	Forskel Moms IP omsætning	Lagerændringer
En faglig enhed	0-75	34,7	8,7	2,3	8,0
	75-100	7,9	3,4	0,0	3,4
	100-200	16,6	9,1	2,3	8,5
	>200	25,4	49,0	20,9	62,3
Flere faglige enheder		15,4	29,8	74,6	14,2
Total		100	100	100	100

Hvad der nok ikke er særligt overraskende, at det er de store og de mere komplekse virksomheder, der tegner sig for størstedelen af omsætningen og hovedparten af den observerede forskel mellem moms og IP omsætning.

Simulation med tre cut-off points

For at se, hvor meget indekset konkret bliver påvirket af anvendelsen af momsdata for 'mindre' virksomheder med én faglig enhed har vi genberegnet indekset for årene 2008 og 2009 med data, hvor vi for en del af stikprøven har brugt momsdata i stedet for vores oprindelige spørgeskemadata. Vi har lavet beregningerne med tre forskellige cut-off points.

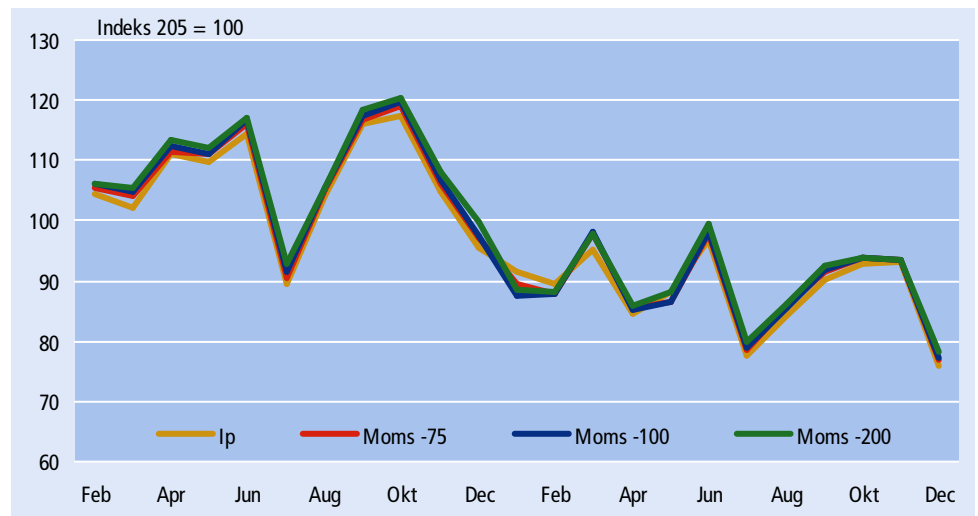
- Momsdata anvendes for alle virksomheder med årsomsætning < 75 mio. kr.
- Momsdata anvendes for alle virksomheder med årsomsætning < 100 mio. kr.
- Momsdata anvendes for alle virksomheder med årsomsætning < 200 mio. kr.

Lager sat lig nul

Da vi ikke har noget alternativ til lagerdata, har vi valgt at sætte lagerværdien lig nul for alle de virksomheder, hvor vi bruger moms data. Dette udgør selvfølgelig et kvalitetsproblem, og det er muligt, at det vil være nødvendigt at komme frem til en mere holdbar løsning, hvis man skulle anvende moms data i produktion.

Industriproduktionsindekset 2008 – 2009

Figur 1



Sammenligning af indeks

Som man kan se følger de fire indeks hinanden meget tæt, men der er dog enkelte punkter hvor der er en pæn forskel mellem de rene spørgeskemabaserede data og de data, der er beriget med momstal. Primært i begyndelsen af 2009, hvor udsvingene er lidt større i de momsberigede data. IP falder med 4,1 pct fra december til januar og stiger med 5,9 pct fra februar til marts. De tilsvarende tal for de momsberigede data er et fald på omkring 8 pct. fra december til januar og en stigning på 11 pct. fra februar til marts. Alle tallene der er nævnt her er ikke sæsonkorrigerede, så udsvingene er noget større end de månedstal vi normalt kommenterer på.

Går man ned på et mere detaljeret niveau, viser det sig, at der er markante forskelle inden for nogle branchegrupper, mens de fire serier næsten er identiske inden for andre.

6. Konklusion

Resultaterne af vores undersøgelse er alt i alt opmuntrende. Det ser ud til, at der kan være en mulighed for at undlade at indsamle data for mellem 35 og 60 pct. af stikprøven og stadig nå frem til et resultat, der minder meget om det vi ville få, hvis vi brugte vores indsamlede data for alle enhederne.

Fordele ved brug af moms til IP

Fordelen ville være store, hvis man kunne benytte denne hybride metode, dvs. en blanding af administrative data og spørgeskemadata, til beregning af IP.

For det første ville man kunne reducere respondentbyrden markant for IP ved at bruge momsdata for en stor andel af virksomhederne i stikprøven.

For det andet må man forvente, at der vil være en vis reduktion i arbejdsbyrden for Danmarks Statistik. Der vil dog stadig være en del arbejde forbundet med stikprøvevedligeholdelse, fejlsøgning og datacheck for de virksomheder vi ikke sender skemaer til.

Omkostninger På den anden side vil der også være nogle kvalitetsmæssige omkostninger primært fordi vi pt. ikke har et alternativt bud på lagerændringerne.

Kan vi gøre det billigere og bedre? Som svar på det indledende spørgsmål kan man sige, at vi formodentlig kan producere IP billigere, men det er nok mere tvivlsomt om vi kan gøre det bedre alene ved anvendelse af momsdata. Tværtimod er der risiko for at kvaliteten bliver dårligere med mindre vi finder en god beregningsmetode for lagerændringerne.