

Naar een snellere beschikbaarheid van cijfers

Leendert Hoven

1. Inleiding

Statistiek is in zekere zin bederfelijke waar. Naarmate cijfers sneller beschikbaar komen, zijn ze van meer waarde voor de gebruiker. Met name vanuit Europa wordt steeds meer een beroep gedaan op statistiekproducenten om te komen tot een snellere publicatie van belangrijke korte termijnindicatoren. Het spreekt eigenlijk vanzelf, dat de informatie niet alleen snel moet zijn, maar ook aan andere kwaliteitseisen moet voldoen. Gebruikers die het belang van snellere cijfers bepleiten, noemen vaak in één adem het belang van voldoende nauwkeurigheid. Met name voor beleidsdoeleinden zijn de vereisten qua nauwkeurigheid streng en kunnen de gevolgen van het gebruik van snelle, maar onnauwkeurige cijfers minstens zo ernstig zijn als het laat beschikbaar komen van de cijfers. Snel, maar ook voldoende nauwkeurig, daar gaat het dus om. In dit artikel wordt ingegaan op de vraag, welke mogelijkheden bestaan om tot snellere publicatie te komen zonder dat de nauwkeurigheid van de cijfers in het gedrang komt.

2. Europa en snellere cijfers

De eenwording van Europa heeft geleid tot een groeiende vraag naar betrouwbare, vergelijkbare, maar ook snellere Europese statistische informatie. De roep om snellere cijfers had vooral te maken met het feit dat belangrijke korte termijninformatie in de Verenigde Staten doorgaans veel eerder ter beschikking komt dan in Europa.

Inmiddels is een aantal stappen gezet om aan de vraag naar snellere cijfers te kunnen voldoen.

Allereerst is in 2000 het actieplan van het Economisch en Financieel Comité, EFC-actieplan of EMU-actieplan opgesteld, met als oogmerk om snelle Europese cijfers te kunnen maken voor het economische beleid van Europa. Het plan bestaat uit een aantal aanbevelingen rond versnellingen of uitbreidingen van statistische gegevens die voor de Economische Monetaire Unie (EMU) van belang zijn.

Vervolgens werd in 2001 een task force opgericht die moest inventariseren hoe snel en op welke wijze de korte termijn cijfers in de Verenigde Staten worden samengesteld. Uit deze vergelijking kwam een voorstel voor versnelling van een aantal kernvariabelen, de zogenaamde Principal European Economic Indicators (PEEI's)¹⁾. Voor deze PEEI's zijn vervolgens in Europees verband afspraken gemaakt over versnelling, verdere harmonisering en detaillering.

Ten slotte kan het *First for Europe* principe genoemd worden. *First for Europe* betekent dat cijfers voor Europa

prioriteit hebben boven de nationale cijfers en dat de processen en methoden zo worden ingericht dat er snelle resultaten voor de eurozone beschikbaar komen. Het gaat hierbij in de eerste plaats om de PEEI's.

Verder is van belang dat Europa, en dan met name de Europese Centrale Bank, veel belang hecht aan gelijktijdige publicatie van de cijfers van de lidstaten. Momenteel probeert men, onder andere voor de publicatie van de kwartaalcijfers van de economische groei, tot nadere afspraken hierover te komen. Dat zou concreet kunnen betekenen dat een lidstaat die zijn cijfers eerder klaar heeft dan de andere nog een aantal dagen met publiceren zou moeten wachten.

Het CBS ziet de noodzaak van vollediger en snellere cijfers voor Europa. Het *First for Europe* principe wordt door het CBS op hoofdlijnen als volgt ingevuld:

- Voor de Europese statistieken en variabelen, de variabelen uit het EFC-actieplan en de PEEI's voorop, geldt als doelstelling dat het CBS levert binnen de termijnen die in Europees verband zijn afgesproken ten aanzien van *flash releases* en daarop volgende publicaties.
- In principe worden alleen cijfers aan Europa geleverd die ook nationaal worden gepubliceerd en wordt het nationale cijfer op hetzelfde moment als de EU/EMU totalen gepubliceerd. Alleen bij uitzondering kunnen ook cijfers worden geleverd die nationaal later, of niet worden gepubliceerd.

3. Hoe snel is Nederland in vergelijking met andere landen?

Zoals gezegd zijn de Verenigde Staten in het algemeen veel sneller met het publiceren van korte termijnstatistieken dan Europa. In de Verenigde Staten worden belangrijke korte termijnstatistieken al binnen 2 à 3 weken na afloop van de verslagperiode gepubliceerd. Binnen Europa zijn de verschillen qua snelheid aanzienlijk. Alleen het Verenigd Koninkrijk komt voor sommige statistieken (met name de flashraming van de economische groei) in de buurt van de Verenigde Staten. Nederland neemt een middenpositie in: relatief snel op sommige terreinen, minder snel op andere. Tabel 1 vergelijkt de datum van eerste publicatie (gemeten in aantal dagen na afloop van de verslagperiode) voor enkele belangrijke korte termijnindicatoren tussen een aantal Europese landen.

Overigens is enige voorzichtigheid bij de interpretatie van de publicatiedata op zijn plaats. Dat geldt onder meer voor de werkloosheidscijfers. In Nederland worden deze berekend als voortschrijdend driemaandsgemiddelde. Het driemaandsgemiddelde wordt toegekend aan de middelste van de drie maanden waarover wordt gemiddeld. In andere landen waar met driemaandsgemiddelden wordt gewerkt

wordt het cijfer toegekend aan de laatste van de drie maanden. Men is op die wijze ogenschijnlijk een maand sneller, maar in werkelijkheid niet. Een soortgelijke vorm van 'optisch bedrog' doet zich voor bij de inflatiemeting en de producentenprijsindexcijfers, waar de timing en de lengte van de meetperiode binnen de maand van land tot land verschilt. Ten slotte dient vermeld te worden dat de tabel niets zegt over andere kwaliteitsaspecten, zoals nauwkeurigheid en mate van detail. Als men een hoge prioriteit toekent aan snelheid, kan dat soms betekenen dat men genoeg neemt met een wat minder nauwkeurig cijfer. In dit verband wordt er soms op gewezen dat de Amerikaanse cijfers naderhand nog wel eens fors worden herzien. Zo werd bijvoorbeeld het groeicijfer over het eerste kwartaal van 2001, dat binnen een maand na afloop van het kwartaal werd gepubliceerd, enkele weken later alweer fors bijgesteld (1,3 in plaats van 2 procent). Overigens zijn bijstellingen van de groeicijfers in Nederland ook geen uitzondering. Interessant in dit verband is het onderzoek van Nijmeijer en Hijman (2004). Zij constateren dat gemiddeld over het tijdvak 1991–2000 de eerste raming van de economische groei 0,35 procentpunt onder het uiteindelijke cijfer ligt, met een uitschieter van 0,8 procentpunt in 1994.

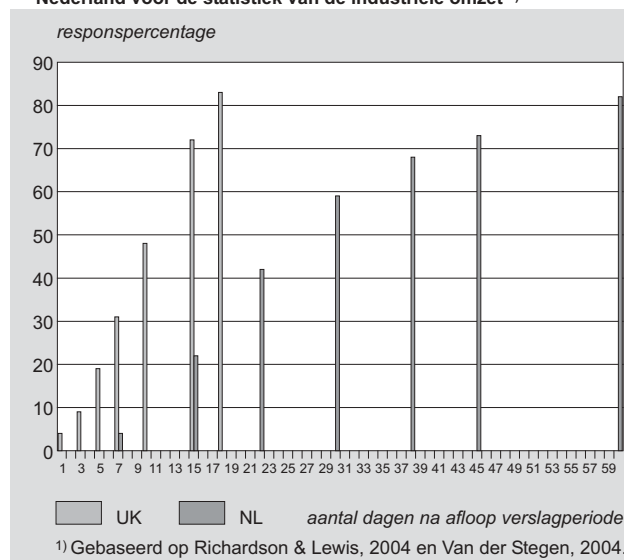
4. Waarom zijn anderen soms sneller dan wij?

Wat betreft de vergelijking tussen Europa en de Verenigde Staten, spelen de volgende aspecten een rol:

- politieke en institutionele verschillen;
- verschillen in methoden, expertise, middelen en prioriteiten.

In de Verenigde Staten wordt een hoge prioriteit aan snelheid gegeven. Deze is formeel vastgelegd in presidentiële richtlijnen, die inhouden dat belangrijke korte termijnstatistieken binnen 22 werkdagen na afloop van de verslagperiode moeten worden gepubliceerd. Aan andere kwaliteitseisen, met name de samenhang tussen statistieken en regionale detaillering, wordt minder belang gehecht. Dat heeft mede te maken met de manier waarop de statistiekproductie in de Verenigde Staten is geregeld: gecentraliseerd op federaal niveau, maar gedecentraliseerd als het om verschillende statistieken gaat. Er is nauwelijks coördinatie tussen de verschillende bureaus (Bureau of

1. Vergelijking van de responsnelheid in het Verenigd Koninkrijk en Nederland voor de statistiek van de industriële omzet ¹⁾



Economic Analysis, Bureau of Labor Statistics, Bureau of the Census, Federal Reserve Board). De eerste ramingen zijn in de VS vaak gebaseerd op kleine steekproeven, maandindicatoren zijn vaak gebaseerd op een bepaalde dag of week in het midden van de maand. Verder steekt men erg veel energie in het snel binnenhalen van de gegevens (*response chasing*). Het aantal vragen is doorgaans beperkt, men maakt veel gebruik van *quick media* (e-mail, fax) en er wordt ook erg veel gebruik gemaakt van de telefoon om de informatie snel binnen te halen.

In Europa ligt de focus traditioneel op nationale statistieken en hecht men meer aan nauwkeurigheid dan aan snelheid. Wat betreft de vergelijking binnen Europa kan worden opgemerkt dat ook het Verenigd Koninkrijk erg veel energie steekt in *response chasing*. Het verschil in responsnelheid tussen het Verenigd Koninkrijk en Nederland is soms opmerkelijk, zie figuur 1 voor een voorbeeld.

5. De afruil tussen snelheid en nauwkeurigheid

Een telkens terugkerend aspect in de discussies over snelle cijfers betreft de afruil tussen snelheid en nauwkeurigheid. Het gaat erom een optimaal evenwicht tussen

Tabel 1
Vergelijking van publicatiedata tussen Europese landen voor enkele korte termijnindicatoren, naar de situatie in de tweede helft van 2003

Economische groei	Inflatie ¹⁾	Producentenprijsindex	Detailhandelsomzet	Productieindex	Bouwproductie	Werkloosheid
UK 25	BE 3	LU 9	UK 22	FI 30	DE 38	AT 3
NL 45	NL 7	UK 10	SE 26	PT 31	FR 43	BE 3
DE 45	LU 7	FI 17	ES 30	ES 37	BE 44	DE 6
FR 47	DK 10	DE 19	PT 31	DK 37	NL ²⁾ 45	IE 7
ES 57	DE 11	PT 20	DE 33	DE 38	PT 46	ES 8
BE 57	FR 12	BE 22	FR 35	UK 39	FI ²⁾ 50	SE 12
DK 60	SE 13	IE 25	DK 40	NL 40	DK ²⁾ 60	PT 16
SE 65	IE 13	ES 26	FI 46	FR 43	LU 64	FI 18
FI 70	ES 13	SE 26	NL 51	IT 43	UK ²⁾ 66	LU 24
PT 70	FI 14	FR 27	IE 52	BE 44		DK 28
IT 71	IT 15	IT 30	IT 52	SE 45		FR 30
GR 71	PT 15	NL 30	BE 57	IE 52		NL 48
AT 80	UK 18	DK 36	AT 58	LU 68		UK 52
IE 110	AT 18			AT 119		IT 55

¹⁾ Dit betreft de 'officiële cijfers', dus niet de flashramingen die door Duitsland en Italië worden geproduceerd en rond het einde van de verslagmaand beschikbaar komen.
²⁾ Kwartaalstatistiek.

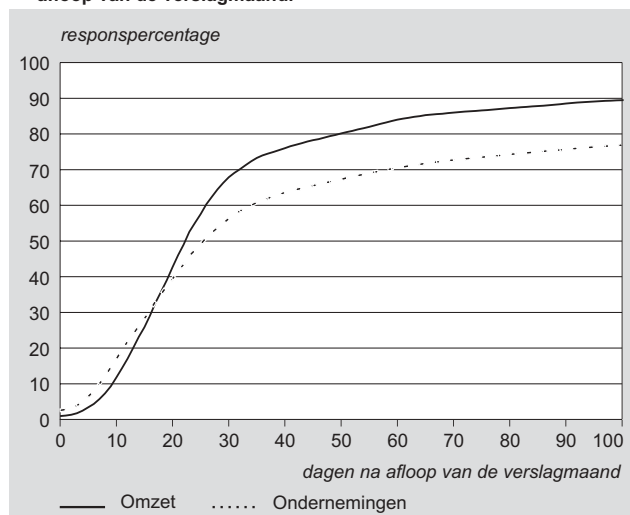
deze kwaliteitsaspecten te vinden. Een probleem daarbij is dat dit evenwicht niet voor elk gebruik, elke gebruiker of voor elk statistisch gegeven hetzelfde zal zijn. Ook zullen de specifieke omstandigheden hierbij een rol kunnen spelen. Op het moment dat de economie zich bijvoorbeeld op een keerpunt bevindt, kan een kleine herziening groei in recessie veranderen of vice versa.

Wat moet precies onder nauwkeurigheid worden verstaan? In de discussies over snellere cijfers hanteert men vaak de volgende maatstaf voor onnauwkeurigheid: de discrepantie tussen een eerder gepubliceerd, voorlopig cijfer en het definitieve cijfer. Soms duidt men deze maatstaf aan met de term 'betrouwbaarheid' om daarmee te onderstrepen dat ook het definitieve cijfer nog kan afwijken van het onbekende 'werkelijke' cijfer.

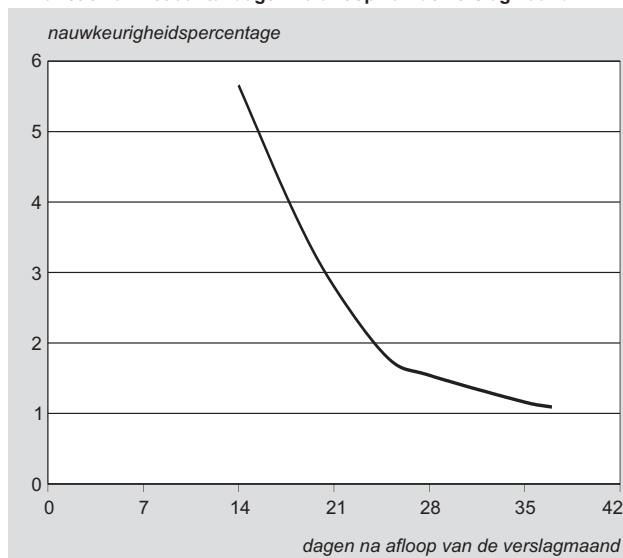
Normaliter zal een cijfer, naarmate het eerder gepubliceerd wordt, minder nauwkeurig zijn omdat het gebaseerd is op minder gegevens. Wordt de informatie bijvoorbeeld verkregen uit schriftelijke enquëtering van (een steekproef van) bedrijven, dan zal niet ieder bedrijf even snel responderen. Figuur 2 (ontleend aan Aelen, 2004) illustreert dit voor de maandstatistiek van de industriële omzet. De figuur laat zien hoe het percentage van het aantal ondernemingen dat gerepondeerd heeft zich in de tijd ontwikkelt, alsmede het percentage van de omzet dat met deze ondernemingen correspondeert. De eerste omzetcijfers die het CBS iedere maand publiceert, 37 dagen na afloop van de verslagperiode, hebben gemiddeld betrekking op de respons van 62 procent van de bedrijven, gezamenlijk goed voor gemiddeld 74 procent van de totale omzet. Voor de bedrijven waarvan op dat moment nog geen respons binnen is wordt een omzetwaarde geraamd ('geïmputeerd') op basis van de veronderstelling dat de ontwikkeling van de omzet van deze bedrijven ten opzichte van de vorige maand gelijk is aan die van overeenkomstige bedrijven die wel hebben gerepondeerd. Men maakt hierbij onderscheid in 124 categorieën bedrijven.

In principe kan op ieder moment na afloop van de verslagperiode, zodra er respons binnen is, een omzetcijfer worden berekend, dat naar mate de tijd voortschrijdt op steeds meer respons gebaseerd is en steeds nauwkeuriger wordt.

2. Ondernemingen die hebben gerepondeerd als percentage van de totale steekproef en de omzet van deze bedrijven als percentage van de totale omzet, beide als functie van het aantal dagen na afloop van de verslagmaand.



3. Nauwkeurigheid van de statistiek van de industriële omzet als functie van het aantal dagen na afloop van de verslagmaand



Dit wordt geïllustreerd in figuur 3. Deze laat de ontwikkeling van de *root mean square error* (rmse) van de voorlopige schatting ten opzichte van het definitieve cijfer zien als functie van het aantal dagen na afloop van de verslagmaand voor de maandstatistiek van de industriële omzet. De figuur is gebaseerd op een simulatie van Aelen (2004), die hiervoor de beschikking had over microdata over een periode van vijf jaar. De rmse is een veel gebruikte maat voor nauwkeurigheid. Andere maatstaven die men zou kunnen gebruiken zijn de maximale fout, het aantal malen dat de fout een bepaalde waarde overschrijdt of het aantal malen dat de richting van de ontwikkeling verkeerd wordt aangegeven. Het cijfer dat momenteel na 37 dagen wordt gepubliceerd heeft een rmse van circa 1,1%. Uit de figuur valt af te lezen hoe groot de nauwkeurigheid zou zijn bij eerdere publicatie, gesteld dat aan de methode verder niets zou worden veranderd en de respons op dezelfde wijze zou blijven binnenkomen als nu het geval is. Zou het streven zijn om al na 15 dagen te publiceren (in navolging van de VS) dan zou de rmse ceteris paribus rond 6 procent uitkomen en daarmee de kans op forse herzieningen toenemen. De vraag is of deze nauwkeurigheid voor gebruikers voldoende zou zijn.

Wanneer is een cijfer nauwkeurig genoeg? Bier en Ahnert (2001) geven vanuit de optiek van de Europese Centrale Bank enkele indicatieve richtlijnen. Voor sleutelmaatstaven, zoals de geharmoniseerde index van consumentenprijzen als maatstaf voor prijsstabiliteit, stelt men dat frequente herzieningen van meer dan 0,1 procentpunt even schadelijk zijn als late publicatie. Voor een reeks andere belangrijke indicatoren die gewoonlijk geringe veranderingen van periode op periode laten zien, zijn aanvaardbare herzieningen vergelijkbaar klein. Voor een uitgebreide reeks conjunctuurgegevens liggen aanvaardbare herzieningen tussen 0,1 en 0,5 procentpunt, afhankelijk van de aard van de indicator. Ten slotte zijn er nog statistische gegevens waarvoor de eisen minder streng zijn. Met name gaat het daarbij om detailinformatie of om gegevens waarvan bekend is dat ze van periode op periode sterk fluctueren.

Het voorgaande betekent niet dat er geen mogelijkheden zouden zijn om sneller te publiceren zonder dat de nauwkeurigheid in het gedrang komt. Het komt er dan op neer om mogelijkheden te zoeken om een lijn als in figuur 3 naar links te verplaatsen. Dat zou bijvoorbeeld kunnen door bedrijven ertoe aan te sporen sneller in te zenden (*response chasing*, zoals in Engeland gebeurt). Maar er zijn ook andere mogelijkheden.

6. Mogelijkheden tot versnellen

Het zou in het kader van dit artikel te ver voeren om alle mogelijkheden om te versnellen de revue te laten passeren. Zonder volledig te willen zijn, noemen we een aantal die veelbelovend lijken of in ieder geval de moeite waard om nader te onderzoeken.

6.1 Gebruik van alternatieve imputatiemethoden voor non-respons

Onderzoek van Aelen (2004) heeft aangetoond dat voor de statistiek van de industriële omzet door het gebruik van alternatieve imputatiemethoden voor non-respons een beter resultaat kan worden behaald. De eerste methode die Aelen heeft onderzocht en die hij de *improved trend-based* methode noemt, lijkt veel op de momenteel gebruikte methode. Er worden echter geen gegevens van de verslagmaand en de maand voorafgaand aan de verslagmaand gebruikt, maar gegevens over de verslagmaand en de overeenkomstige maand één jaar eerder. Daarnaast worden bij het imputeren minder categorieën van bedrijven onderscheiden (30 in plaats van 124).

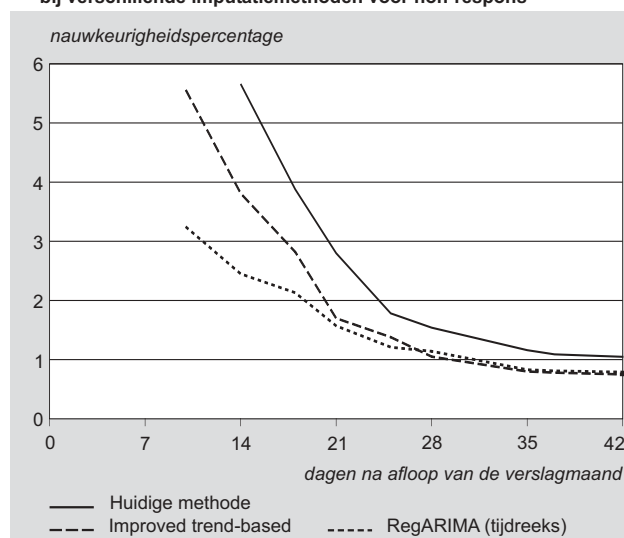
De tweede methode die Aelen heeft onderzocht is gebaseerd op tijdreeksanalyse. Zowel in de huidige methode als in de *improved trend-based* methode worden bij het imputeren de gegevens van slechts twee maanden gebruikt. De regARIMA methode maakt gebruik van de gehele tijdreeks. Hiertoe wordt de populatie van bedrijven in twee groepen verdeeld: een groep die gerespondeerd heeft (groep I) en een groep die dat (nog) niet heeft gedaan (groep II). De tijdreeksen van beide groepen (die van groep I loopt tot maand t , die van groep II tot maand $t-1$) worden vervolgens gebruikt om de totale omzet van groep II voor maand t te voorspellen. Dat gebeurt door de tijdreeks van groep II te modelleren volgens de zogenaamde regARIMA methode:

$$Y_{it} = \beta_0 + \beta_1 Y_{it} + Z_t,$$

waarin Y de totale omzet aangeeft, de subscripten I en II de genoemde groepen aangeven, β_0 en β_1 regressiecoëfficiënten zijn en Z_t de residuen. De residuen worden gemiddeld met gebruik van ARIMA modellen, die vaak gebruikt worden bij de analyse van economische tijdreeksen. Voor verdere details wordt verwezen naar Aelen (2004).

Figuur 4, ontleend aan Aelen (2004), laat de resultaten van beide alternatieve imputatiemethoden in vergelijking met de huidige methode zien. Beide methoden presteren aanmerkelijk beter, in termen van de rmse. Beide methoden leveren rond 27 dagen na afloop van de verslagmaand, wanneer gemiddeld 52 procent van de bedrijven heeft gerespondeerd, een resultaat dat net zo nauwkeurig is als

4. Nauwkeurigheid van de statistiek van de industriële omzet als functie van het aantal dagen na afloop van de verslagperiode, bij verschillende imputatiemethoden voor non-respons



het momenteel na 37 dagen gepubliceerde cijfer (dat berust op 62 procent respons). De figuur laat zien dat beide alternatieve methoden ook bij eerdere publicatie beter presteren dan de huidige methode. Vanaf 3 weken na afloop van de verslagperiode presteren de *improved trend-based* methode en de regARIMA methode ongeveer even goed, bij nog eerdere publicatie is de regARIMA methode beter.

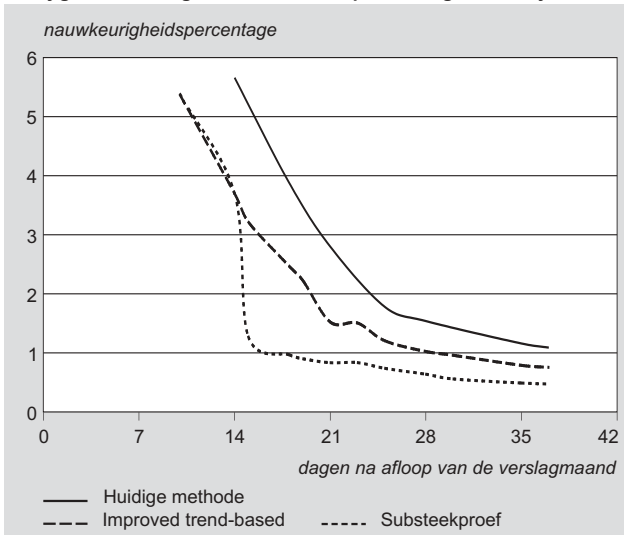
6.2 Gebruik van substeekproeven met de belangrijkste bedrijven

Een andere methode die veelbelovend lijkt maakt gebruik van een substeekproef waarin alleen de belangrijkste bedrijven zitten. Van der Stegen (2004) heeft deze methode onderzocht voor de maandstatistiek van de industriële omzet. Van der Stegen is in zijn onderzoek uitgegaan van een substeekproef van bedrijven met 500 of meer werknemers. Deze bedrijven, ongeveer 3 procent van het totale aantal, zijn goed voor meer dan 40 procent van de industriële omzet. Als het mogelijk zou zijn om de respons van deze bedrijven (via gerichte *response chasing*) al snel na afloop van de verslagperiode binnen te halen, dan kan al binnen 2 à 3 weken een resultaat worden behaald dat net zo nauwkeurig is als momenteel na 37 dagen. De gegevens van de andere bedrijven worden, voor zover ze binnen zijn, ook meegenomen, maar er wordt geen moeite gedaan om deze versneld binnen te krijgen. Door deze methode te combineren met een betere imputatiemethode voor non-respons, zoals de *improved trend-based* methode, kan het resultaat nog verder worden verbeterd. Zie figuur 5, waarin is aangenomen dat alle grote bedrijven binnen twee weken hebben gerespondeerd.

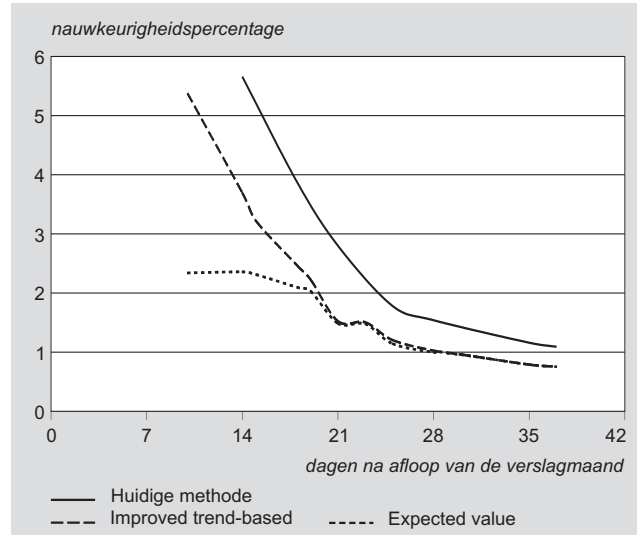
6.3 Nowcasten/Expected Value Approach

Nowcasten is afgeleid van *forecast* (voorspellen) en betekent zoiets als 'het voorspellen van het heden'. Volledig *nowcasten* houdt in dat nog geen waarnemingen van de

5. Nauwkeurigheid van de statistiek van de industriële omzet als functie van het aantal dagen na afloop van de verslagperiode, bij gebruikmaking van een substeekproef van grote bedrijven



6. Nauwkeurigheid van de statistiek van de industriële omzet als functie van het aantal dagen na afloop van de verslagperiode, bij gebruikmaking van de expected value approach.



doelvariabele binnen zijn en dat een raming wordt gemaakt op basis van een model. Daarin kan gebruik gemaakt worden van tijdreeksanalyse, verwachtingsindicatoren (bijvoorbeeld conjunctuurtestgegevens) of gerelateerde indicatoren (bijvoorbeeld het elektriciteitsverbruik als indicator voor de industriële productie).

De *expected value approach* combineert de *nowcasts* met vroeg binnengekomen respons (die op zijn beurt weer gebruikt kan worden om de kwaliteit van de *nowcasts* te verbeteren). De aldus tot stand gekomen ramingen combineren waarnemen (meten) met modelmatig schatten en krijgen, naar mate er meer waarnemingen binnenkomen, steeds meer het karakter van een echte statistische raming.

Een eerste studie naar de toepasbaarheid van de *expected value approach* heeft een aantal interessante resultaten opgeleverd (zie Van der Stegen, 2004, die dit heeft onderzocht voor de maandstatistiek van de industriële omzet). Gedurende de eerste twee weken na afloop van de verslagmaand, als nog heel weinig respons binnen is, levert deze benadering in vergelijking met andere methoden (alternatieve imputatiemethoden, substeekproeven) veruit de beste resultaten op (zie figuur 6). De nauwkeurigheid van de beginschatting lijkt echter nog onvoldoende voor publicatie. Daarnaast moet worden opgemerkt dat de combinatie van beginschatting met vroege respons gedurende de eerste twee à drie weken niet tot een aanwijsbare verbetering ten opzichte van de beginschatting leidt. Dit wordt veroorzaakt door de grote variantie van de vroege respons. Uit vervolgonderzoek zal moeten blijken of hiervoor een oplossing gevonden kan worden.

6.4 Keuze van geschikte (vroege) peilperioden

Deze methode wordt vooral toegepast in de Verenigde Staten en is in kringen van statistici niet geheel onomstreden. Sommigen zijn van mening dat het kiezen van een vroeg peilmoment alleen de definitie verandert en helemaal niets versnelt.

In Europa wordt deze aanpak eigenlijk alleen toegepast bij prijsstatistieken. In een aantal gevallen is er daarbij expliciet voor gekozen om het prijsniveau op een specifiek moment (bijvoorbeeld een bepaalde dag van de maand) te vergelijken met een ander specifiek moment (dezelfde dag in een andere maand) en zijn de resulterende cijfers nadrukkelijk niet bedoeld als benadering van het gemiddelde prijsniveau over de gehele maand. In dat geval kan men dus niet stellen dat een land dat bijvoorbeeld de prijsniveaus van de 1e van iedere maand vergelijkt, sneller zou zijn dan een land dat zich ten doel stelt om de gehele maand te representeren. De verslagperiode verschilt immers.

Anders wordt het als men een subperiode binnen de maand kiest met de bedoeling om de gehele maand te representeren. In dat geval is de vraag aan de orde in hoeverre men daarin slaagt. Het is duidelijk dat als men bijvoorbeeld de prijzen gedurende de eerste twee weken van de maand meet, een prijsexplosie van benzine of een prijsenorlog tussen supermarkten die in de tweede helft van de maand uitbreekt niet in het cijfer van de desbetreffende maand tot uitdrukking komt. Van Ruth (2002) heeft laten zien dat de onnauwkeurigheid van de cijfers voor producten met frequent variërende prijzen snel toeneemt naarmate men gedurende een kortere periode meet. Ook hier is dus weer de afruil tussen snelheid en nauwkeurigheid aan de orde. Een opmerkelijke uitkomst van het onderzoek was dat de timing binnen de maand van weinig invloed was op de nauwkeurigheid. Een week in het midden van de maand was in de onderzochte periode niet nauwkeuriger dan de eerste week van de maand.

6.5 Verbeteringen op het terrein van de waarneming

Hier valt waarschijnlijk ook de nodige winst te behalen. Te denken valt onder meer aan beperking van de enquête-ting, bijvoorbeeld door waar mogelijk gebruik te maken van bestaande registers en administraties (op zich komt de informatie uit deze bronnen overigens vaak eerder langzamer dan sneller ter beschikking), optimaliseren van de

doeltreffendheid van de enquêtering (betere vragenlijsten), maximaal aansluiten bij (bedrijfs-)administratieve en fiscale concepten, verdere professionalisering van de (r)appelstrategie en alternatieve vormen van dataverzameling. Verbeteringen op dit terrein kunnen er mogelijk ook toe leiden dat de kwaliteit van de verzamelde gegevens beter is, waardoor vervolgens ook weer minder tijd nodig is voor controle en editing.

7. Conclusies

De behoefte aan snelle statistische informatie groeit steeds verder. Het CBS erkent het belang van snelle cijfers en onderzoekt momenteel hoe op verantwoorde wijze invulling aan deze behoefte kan worden gegeven. De resultaten van het onderzoek dat tot dusverre is verricht zijn bemoedigend. Er zijn verschillende mogelijkheden om te versnellen zonder dat de nauwkeurigheid van de cijfers te zeer in het gedrang komt. Door gebruik te maken van betere imputatiemethoden voor non-respons of van substeekproeven waarin de belangrijkste bedrijven zitten, waarvan men de gegevens vervolgens snel binnenhaalt, kan bij de maandstatistiek van de industriële omzet al een flinke tijdswinst worden geboekt. Daarnaast kan mogelijk ook het gebruik van modelmatige schattingsmethoden snelle eerste ramingen van voldoende nauwkeurigheid opleveren. Hiernaar is nog verder onderzoek nodig. Ten slotte valt ook op het terrein van de waarneming nog winst te behalen.

Literatuur

- Aelen, F., 2004, Improving timeliness of industrial short-term statistics using time series analysis, CBS Discussion paper 04005, Centraal Bureau voor de Statistiek, april 2004.
- Bier, W. & H. Ahnert, 2001, De afruil tussen tijdigheid en nauwkeurigheid, Economisch Statistische Berichten, 86e jaargang, nr. 4299, 15 maart 2001.
- Nijmeijer, H. & R. Hijman, Bijstellingen ramingen economische groei, De Nederlandse conjunctuur, 2004, aflevering 1, Centraal Bureau voor de Statistiek.
- Öberg, S. en Nanopoulos, P. (2001), Report of the task force on benchmarking in infra-annual economic statistics to the SPC. Eurostat, Luxemburg.
- Richardson, I. & A. Lewis, 2004, UK working practices that allow an early delivery of the production index.
- Ruth, F.J. van, 2002, The influence of the measurement period on consumer price indices, CBS rapport, Centraal Bureau voor de Statistiek.
- Stegen, R.H.M. van der, 2004, Improved quality of statistics of manufacturing turnover: timeliness and accuracy, CBS rapport (nog te verschijnen), Centraal Bureau voor de Statistiek.

Noot in de tekst

- 1) Zie Öberg & Nanopoulos, 2001.