

Hoofdstuk 13
De nauwkeurigheid
van geïntegreerde gegevens



*Peter Kooiman
Kees Zeelenberg
Wouter J. Keller*

1. Inleiding

Herlezing van de bijdrage van Henk van Tuinen aan Denken en Meten wekt associaties met de romans van Jules Verne. Visionair, op sommige punten getuigend van een verbluffend vooruitziende blik, op andere onvermijdelijk de plank misslaand. De automatisering begon goed op gang te komen en het statistisch proces te doordringen. Maar wie kon met enige precisie de stormachtige verdere ontwikkeling van de automatisering op dat moment overzien, laat staan de effecten die ze zou hebben op het statistiekproces. Automatiseren was een stuk lastiger dan verwacht: APS (Automatiseringssysteem Productiestructuur) bleek een brug te ver. Ook moet de *dealing room* van statistische specialisten, die Henk van Tuinen aan het slot van zijn bijdrage schetst, met de integrator als hoekman die vraag en aanbod van statistische informatie in evenwicht brengt, vooralsnog tot de *science fiction* gerekend worden. Op andere punten ging het veel harder dan de meesten van ons konden voorzien. De integratie van onderop, zoals nu veel meer concreet geworden bij het Sociaal statistisch bestand (SSB) en het Economisch statistisch bestand (ESB), en het automatisch gaafmaken, waar de laatste jaren zo'n grote vooruitgang mee is gemaakt, werden door Henk van Tuinen knap voorzien. Ook besteedt hij ruimschoots aandacht aan de organisatorische veranderingen in het statistiekproces die *bottlenecks* zouden blijken. Ook dat bleek te getuigen van een vooruitziende blik. Zo herinnert een onzer zich een voorstel bij de reorganisatie 1994 om voor alle grote berichtgevers de waarneming voor alle statistieken te centraliseren in een *large business unit* die *overall* consistente cijferbeelden per grote berichtgever (plus een consistente opsplitsing over de onderliggende bedrijfseenheden!) zou verzorgen. Een kanteling *avant la lettre* die het helaas niet haalde. Dat zou de *ex ante* integratieverschillen zonder enige twijfel belangrijk hebben teruggebracht.

In paragraaf 3 van zijn artikel gaat Henk van Tuinen in op het waarheidsgehalte van de geïntegreerde statistische gegevens, en daarmee verbonden het vraagstuk van de bepaling van de nauwkeurigheid. Henk van Tuinen is daarover nogal terughoudend: er zijn veel en grote problemen, de gegevens zijn noodzakelijk vertekend, en het is gezien de complexiteit van het onderliggende dataverwerkingsproces ondoenlijk om een kwantitatieve nauwkeurigheidsmaat aan de uitkomsten te verbinden. In deze bijdrage willen wij met name op deze punten ingaan. In dit verband de volgende anekdote. In de zomer van 2000 werd onder regie van Henk van Tuinen een nieuw meerjarenprogramma opgesteld. Een onzer leverde daarvoor een concept bijdrage over het uit te voeren

strategische onderzoek. Daarin de stellingname: "Massaal gebruik van secundaire bronnen als basis voor statistische uitkomsten roept de vraag op naar de beheersing van de kwaliteit. Weliswaar levert confrontatie van bronnen inzicht in de onderlinge verschillen; de relatie met de werkelijkheid kan uiteindelijk alleen worden vastgesteld door eigen waarneming." Henk van Tuinen reageerde onder andere met: "De stelling dat de relatie met de werkelijkheid uiteindelijk alleen kan worden vastgesteld door eigen waarneming zou ik niet zomaar voor mijn rekening willen nemen. Het is een illusie dat eigen waarneming veelal een zuivere weergave van de werkelijkheid oplevert" (e-mail, 31 juli 2000). De auteur van de stelling reageerde vervolgens: "Ik besef dat ik daar een vrij extreem standpunt inneem, en heb er vrede mee als dat wordt afgezwakt. Maar het is wel een heel principieel punt, namelijk: waarop is uiteindelijk onze claim gebaseerd dat we valide gegevens over de samenleving naar buiten brengen, die door wetenschap en beleid als (onbetwist) uitgangspunt genomen kunnen worden? Een kentheoretisch vraagstuk waarop we een helder antwoord zouden moeten hebben dat de samenleving overtuigt dat onze cijfers ten principale geen artefacten zijn die vooral de vooroordelen van een groep vrijgestelden vertegenwoordigen, of daardoor vertekend zijn. Ik kan daarvoor niks anders verzinnen dan pijnlijk nauwkeurige eigen waarneming. ...Overigens ben ik met jou van mening dat geïntegreerde gegevens doorgaans een veel betrouwbaarder beeld geven dan de niet-geïntegreerde; je voelt dat aan je water. Maar gevoelens in de onderbuik vormen een niet erg overtuigend antwoord op het kentheoretische probleem dat ik boven probeer te schetsen. Ben daarom wel benieuwd naar jouw visie op de vraag wat het uiteindelijke houvast is wat statistici hebben voor de claim dat hun uitkomsten een valide beeld van de werkelijkheid schetsten."

Dat is dus de vraag naar de methodologie van de integratie. Het artikel van Henk van Tuinen biedt hier niet veel houvast: er gaat veel informatie in de NR-mixer, zoveel is duidelijk, maar waarin uiteindelijk de correspondentie met de werkelijkheid gelegen is, blijft duister. Rob Mokken sprak, als het gesprek op de Nationale rekeningen kwam, niet helemaal ten onrechte van Voodoo-methoden. De methodologie van een vak is de systematische reflectie op de normen en waarden die de beoefenaars hanteren, en de wijze waarop deze een waarborg bieden voor zinvolle resultaten. Voor de natuurwetenschappelijke methode is dat uitvoerig beschreven. Wie 's zaterdags de columns van Piet Borst of Vincent Icke in NRC-Handelsblad leest, krijgt het exemplarisch gedemonstreerd. Op de universiteit krijgen studenten les in Popper, Kuhn en Feyerabend. *Anything goes* is de boodschap waar het gaat om de wijze waarop inzichten gevormd worden. Maar in de empirische wetenschappen, die pretenderen kennis over de werkelijkheid buiten ons te verwerven, ligt de uiteindelijke toets in de congruentie van onze inzichten met pijnlijk nauwkeurige experimentele waarneming gericht op mogelijke weerlegging van deze inzichten. Openbaarheid, reproduceerbaarheid en onpartijdigheid zorgen ervoor dat incongruenties tussen theorie en werkelijkheid, hoe klein ook, niet blijvend onder de mat geveegd kunnen worden.

2. Statistische methodologie

Statistiek als wetenschap is " . . . a logic and methodology for the measurement of uncertainty and for an examination of the consequences of that uncertainty in the planning and interpretation of experimentation and observation." (Stigler, 1986, blz.1). Statistiek is dus een methodologie voor het meten van onzekere

grootheden door middel van experimenten en waarneming. Deze ruime omschrijving omvat zowel methoden uit de Neyman-Pearson school van het toetsen van hypothesen, als methoden uit de Bayesiaanse, subjectieve, school, maar ook die uit de daartussen staande Fisheriaanse school van de aannemelijkheid (Edwards, 1972; Hacking, 1965). Alle methoden gaan uit van waarnemingen (data) en geven regels voor het beoordelen van uitspraken over de populatie waarop de waarnemingen betrekking hebben. In de woorden van Fisher (1925, blz. 1): “Statistics may be regarded as (i) the study of populations, (ii) as the study of variation, (iii) as the study of the methods of the reduction of data.” Een gemeenschappelijk onderdeel van alle methoden is het schatten, dat is het verkrijgen van informatie over onbekende grootheden die een rol spelen in het gebied van onderzoek (Fisher, 1956, blz. 146).

Statistische methodologie is algemeen aanvaard in het steekproefonderzoek, in de beschrijving van zowel de opzet van dergelijke onderzoeken als de resultaten ervan, getuige de vele literatuur over methoden voor het trekken en ophogen van steekproeven en voor het ontwerp en de analyse van experimentele proefopzetten, bijvoorbeeld Cochran, 1977, Särndal e.a., 1992, Cochran en Cox, 1950.

Wij stammen alle drie uit de hoek van de statistische methoden. De methodologie van het primaire surveyproces is een goed ontwikkeld en levendig vakgebied, deels verankerd in de wiskundige statistiek, deels in de cognitieve psychologie. Het centrale paradigma is *Total Survey Design* (zie bijvoorbeeld Renssen e.a., 1998). Inventariseer en controleer alle mogelijke foutenbronnen; ontwerp het totale proces zodanig dat de schattingsfout, ofwel de afwijking tussen de uitkomst en de onbekende werkelijke waarde, geminimaliseerd wordt, gegeven het beschikbare budget. Alle regels en richtlijnen waaraan een goed uitgevoerd steekproefsgewijs veldonderzoek moet voldoen, zijn expliciet gemotiveerd vanuit deze probleemstelling.

De schattingsfout is de som van de steekproeffout en de niet-steekproeffouten. De steekproeffout meet de fout die het gevolg is van het feit dat we slechts een steekproef waarnemen en niet de gehele populatie. Als de steekproef gebaseerd is op een correct uitgevoerd kansexperiment, en een geschikte bijpassende schatter wordt gehanteerd, dan is de verdeling van de steekproeffout netjes gecentreerd rond de (onbekende) werkelijke waarde, en kunnen we een steekproefmarge bepalen als statistische maat voor de afwijking tussen de gevonden schatting en de werkelijkheid. De niet-steekproeffouten zijn moeilijker in kaart te brengen. Het methodologische recept is hier vooral: zorgvuldige procedures toepassen die waarborgen dat deze fouten niet kunnen overheersen. Steekproefkaders goed bijhouden, vragenlijsten grondig testen, non-repons bestrijden, uitkomsten valideren door herenquëtering, zorgvuldig editten, herwegen naar bekende populatietotalen van hulpvariabelen.

Dat werkelijke surveyprocessen lang niet altijd aan de strenge regels en richtlijnen van deze methodologische uitgangspunten gehoorzamen, doet niets af aan de validiteit van deze methodologie per se. Niet alleen biedt ze de grondslag voor de legitimiteit van de geldigheidsclaim als de regels wél zorgvuldig worden nageleefd; ook kunnen we vrij aardig kwalitatief, en soms ook kwantitatief, aangeven wat er mis kan gaan als deze regels geschonden worden, dat wil zeggen in welke mate ons

vertrouwen in de juistheid van de uitkomsten vermindert als er de hand mee gelicht wordt. Waar behoefte aan is, is een systematische methodologie van de integratie die op geloofwaardige wijze hetzelfde soort waarborgen biedt. Het APS-project kan wellicht gezien worden als een poging een dergelijke methodologie te implementeren.

De allereerste Nationale rekeningenpublicaties (zie bijvoorbeeld CBS, 1953) gaven kwantitatieve maatstaven voor de nauwkeurigheid van de geïntegreerde gegevens. In latere publicaties werd deze praktijk verlaten. In de afgelopen jaren is weer meer aandacht gekomen voor kwaliteit en nauwkeurigheid in de Nationale rekeningen. Ze zijn echter vooral benaderd vanuit de documentatie van het proces; zie bijvoorbeeld de IMF *Data Quality Standard* (Bloem, 2002) en de audit-onderzoeken van het BNP-Comité van Eurostat (zie bijvoorbeeld Akritidis, 2002, en ONS, 2002, voor de Britse methodebeschrijving). Daarnaast wordt nauwkeurigheid vaak gezien als de omvang van de opeenvolgende revisies van voorlopige naar nader voorlopige naar definitieve raming; zie York en Atkinson (1997), ABS (2002, hoofdstuk 29) en Kazemier en Van Rooijen (2002). Voor inpassingen, met name het aanpassen van een dataverzameling aan bepaalde randvoorwaarden, is de al meer dan 50 jaar oude methode van Stone (Stone e.a., 1942), weer in de aandacht gekomen, zie bijvoorbeeld Wroe e.a. (1999) voor toepassingen.

Elk van deze methoden poogt op meer of minder heuristische wijze de nauwkeurigheid van NR-gegevens te beschrijven. Documentatie, hoewel een noodzakelijke voorwaarde voor kwaliteit, is echter geen voldoende voorwaarde. Kleine revisies zeggen weinig over de kwaliteit van de reeks zelf: in het uiterste geval kunnen revisies zelfs geheel vermeden worden door óf alleen nog maar definitieve cijfers te publiceren óf voorlopige cijfers nooit meer te herzien. De methode van Stone, onder andere verder uitgewerkt door Van der Ploeg (1982), komt nog het dichtst bij een methodologische benadering van het integratieproces, zij het dat de toepassingen ervan meestal beperkt blijven tot de laatste fase in dat proces, het inpassen (zie paragraaf 0, bladzijde 119).

In het navolgende proberen we een verbinding te leggen tussen wat wij van het integratieproces menen te begrijpen en de meer formele wiskundige statistiek. We doen dit door eerst te beschrijven wat integratie inhoudt, hoe integratoren te werk gaan, en dit daarna te vertalen in begrippen en technieken uit de statistische methodologie.

3. Integratie

Integratie als statistische techniek is ontstaan bij de samenstelling van de Nationale rekeningen in de jaren dertig en veertig van de vorige eeuw. Ze had daar als doel aansluiting te verkrijgen tussen waargenomen gegevens en de begrippen van de toen net ontstane macro-economische theorie¹⁾. Meer in het algemeen vindt integratie haar bestaansrecht in de toenemende complexiteit en samenhang van maatschappelijke verschijnselen, die een systeembenadering noodzakelijk maken (Keuning en Zeelenberg, 2002).

Verschillen tussen waargenomen gegevens en concepten kunnen verdeeld worden in 5 groepen (FCSM, 2001):

- Steekproeffouten ontstaan door variabiliteit in de data als ze verzameld zijn door middel van een steekproef. Deze fouten, gemeten door de variantie, kunnen geschat worden, ook voor complexe steekproeven.
- Non-responsfouten ontstaan doordat een berichtgever in het geheel niet respondeert, of doordat hij bepaalde vragen niet beantwoordt. Vaak wordt aangenomen dat deze non-respons niet selectief is, zodat ervoor gecorrigeerd wordt door middel van dezelfde technieken die bij de ophoging van een steekproef gebruikt worden.
- Dekkingsfouten ontstaan doordat het kader dat bij het onderzoek gebruikt is, bijvoorbeeld om een steekproef uit te trekken, niet volledig correspondeert met de populatie waarvoor gegevens verzameld moeten worden. Ontbreken bepaalde populatie-eenheden in het kader, dan is er sprake van onderdekking; komen er andere dan populatie-eenheden in het kader voor, of komen bepaalde populatie-eenheden meer dan éénmaal in het kader voor, dan is er sprake van overdekking.
- Meetfouten zijn het verschil tussen de waargenomen waarde en de echte waarde van een gemeten grootte. Ze kunnen gerelateerd zijn aan de vraagstelling, aan de wijze van onderzoek en aan de wijze van beantwoorden. De vertaling van een bepaald begrip in de vraagstelling kan bijvoorbeeld niet correct zijn, de interviewer kan een vraag onjuist toelichten, of een respondent kan een vraag verkeerd begrijpen.
- Bewerkingsfouten ontstaan tijdens het bewerken van de gegevens en kunnen variëren van typefouten tot onjuiste correcties, bijvoorbeeld door het gebruik van onjuiste gaafmaakmethoden.

Het integratieproces behelst de eliminatie van deze verschillen en verloopt in een aantal stappen (CBS, 1992, blz. 22):²⁾

1. Harmoniseren: verschillen in definities, waarderingen, indelingen en detaillering tussen de verschillende bronnen worden gecorrigeerd. Zo moeten de boekhoudkundige begrippen van bedrijven vertaald worden naar de overeenkomstige NR-begrippen, goederenspecificaties moeten vertaald worden naar de NR-productenindeling, e.d. Een voorbeeld is de loonsom die op drie manieren waargenomen wordt: in de Productiestatistiek (PS) als de loonsom uit de jaarrekening van bedrijven, in de Enquête Werkgelegenheid en Lonen (EWL) als loonsom uit de loonadministratie en in de Statistiek Verloonde Dagen als loonsom voor de sociale verzekering.
2. Completeren: populatieverschillen worden gecorrigeerd: overdekking wordt geëlimineerd en voor onderdekking worden bijschattingen gedaan. Eén van de aanpassingen voor onvolledigheid is die voor zwarte activiteiten, waaronder fraude (OESO 2002). Er is een groot aantal methoden voor completering, die meestal gebaseerd zijn op enkelvoudige verhoudingen, zoals extrapolaties van een waarde uit een eerder jaar met behulp van een loonindex, of verhogen van een bedrag met een percentage dat aan een externe bron, bijvoorbeeld een rechtszaak, ontleend is.
3. Minimaliseren van meetfouten: meetfouten, met name de niet-steekproeffouten worden geëlimineerd, onder andere door meerdere bronnen met elkaar te vergelijken en door definitie-vergelijkingen te gebruiken. Bij de raming van de consumptieve bestedingen van gezinshuishoudingen worden bijvoorbeeld drie bronnen met elkaar vergeleken: de Productiestatistiek Detailhandel, het Budgetonderzoek en externe onderzoeken uit vakbladen.

4. Inpassen, het eigenlijke integreren: relevante identiteiten worden aan de data opgelegd, plausibiliteitscontroles worden uitgevoerd en correcties op grond hiervan doorgevoerd, en laatste kleine verschillen worden weggewerkt, meestal door middel van een statistische techniek als RAS. Plausibiliteitscontroles zijn meestal gebaseerd op analyses van de ontwikkeling ten opzichte van eerdere jaren, zowel van de waarde, de volume-index en de prijsindex, en op analyses van de productiviteitsontwikkeling.

Benadrukt moet overigens worden dat, ondanks de verschillen tussen gegevens en concepten van de basisstatistieken en die van NR, het overgrote deel van de NR gebaseerd is op directe waarneming.

De hierboven gegeven beschrijving van het integratieproces kan als volgt samengevat worden in termen van de statistische methodologie. Uitgangspunt is een aantal databronnen, de basisstatistieken, die door de publieke statistiek door directe waarneming verzameld zijn, vaak met het oogmerk te dienen als basis voor de NR. Deze basisgegevens worden getransformeerd tot NR-gegevens met behulp van externe gegevens en modellen voor ontbrekende data, rekening houdend met de betrouwbaarheid van basis- en externe gegevens. Een deel van de transformatie kan deterministisch zijn, als basisgegevens 1-op-1 overgenomen worden in de NR, zoals bijvoorbeeld gebeurt met gegevens van belastingopbrengsten, of als basisgegevens met een door een externe bron gegevens percentage verhoogd worden. Vaker is de transformatie echter stochastisch: de basisgegevens worden dan getransformeerd rekening houdend met hun betrouwbaarheid. Zo'n stochastische transformatie gebeurt lang niet altijd expliciet: soms worden op grond van plausibiliteitsanalyses correcties aangebracht zonder dat de argumentatie hiervoor duidelijk gedocumenteerd wordt.

4. Integratie en statistische methodologie

Het doel van integratie is het schatten van een samenhangend stelsel van populatiegrootheden door het combineren van meerdere bestaande bronnen. Dat bestempelt de integratie tot een vorm van secundaire data-analyse. Hoewel integratoren soms terug kunnen grijpen op onderliggend micro-materiaal is de meer typische situatie die waarbij de integratie vertrekt van schattingen van populatiegrootheden op meso-niveau, die zelf gebaseerd zijn op steekproefsgewijs onderzoek, op administratieve bronnen, of op combinaties van beide. De kwaliteit van de voor de integratie beschikbare bronnen is voor de integratoren een gegeven, ze hebben daar geen directe invloed op.

In relatie tot de doelparameters van de integratie, dat wil zeggen, de populatiegrootheden die men statistisch wil beschrijven, zijn deze bronnen in de praktijk veelal behept met aanzienlijke niet-steekproeffouten, dat wil zeggen, ze zijn sterk vertekend. Zoals in de vorige paragraaf reeds naar voren kwam, kan het grootste deel van het integratieproces daarom beschreven worden als een poging om deze vertekening te reduceren door bijschatting van ontbrekende delen en andere correcties. Daarbij wordt impliciet of expliciet de vertekening in de voor integratie beschikbare bronnen statistisch gemodelleerd. Dit laatste biedt het aangrijpingspunt voor het ontwikkelen van een meer formele methodologie van het integratieproces. We beschrijven onze ideeën voor een meer statistische benadering van het integratieproces aan de hand van een voorbeeld.

Om voor ons doel niet essentiële complicaties te vermijden, bekijken we het meest simpele voorbeeld van integratie dat wij ons kunnen voorstellen: twee bronnen voor één populatiegrootte:

1. Bron 1: schatting t_1 voor grootte θ_1 met vertekening v_1 en steekproeffout ε_1 .
2. Bron 2: schatting t_2 voor grootte θ_2 met vertekening v_2 en steekproeffout ε_2 .
3. Balansrestrictie: $\theta_1 = \theta_2 (= \theta)$.
4. Vertekening: $v_i = c_i + \mu_i; i = 1, 2$.

De balansrestrictie is hier wel erg simpel: ze geeft aan dat de beide bronnen gebruikt worden voor de schatting van een en dezelfde grootte δ . De vertekening in de beide beschikbare bronnen is gemodelleerd als een constante c_i (zeg de bijschatting) plus een fout μ_i die de onzekerheid in deze bijschatting vertegenwoordigt. De steekproeffouten ε_i en de bijschattingsfouten μ_i beschouwen we als kansvariabelen met verwachting nul; het lijkt voorts niet onredelijk om te veronderstellen dat ze tussen bronnen onafhankelijk zijn. Door de diverse veronderstellingen samen te voegen krijgen we:

$$t_1 = \delta + c_1 + \eta_1$$

$$t_2 = \delta + c_2 + \eta_2$$

waarbij de fout η staat voor $\varepsilon + \mu$.

Als we veronderstellen dat de bijschattingen c_i bekend zijn (dat is in onze opzet het werk van de integratoren), dan is hier sprake van een zeer eenvoudig regressiemodel met slechts twee waarnemingen waarbij we de constante term willen schatten. Om dat goed te kunnen doen, moeten we iets weten van de beide storingen. Deze vertegenwoordigen de onzekerheid in onze twee bronnen. Deze onzekerheid is de som van een 'objectieve' component, namelijk de steekproeffout ε , en een 'subjectieve' component, namelijk de onzekerheid μ in onze schatting c van de vertekening v . Voor een schatting van δ is alleen de verhouding van de varianties van η_1 en η_2 van belang: de optimale schatting wordt gevonden als een met deze varianties gewogen gemiddelde van $(t_1 - c_1)$ en $(t_2 - c_2)$. Dit is een plausibel resultaat dat heel dicht aan ligt tegen wat integratoren natuurlijkerwijs geneigd zijn te doen: de beschikbare bronnen eerst zo goed mogelijk zuiveren voor de vertekening in die bronnen, en vervolgens bij elkaar wegen op zodanige wijze dat de meest betrouwbare schatting het meeste gewicht krijgt in het uiteindelijke resultaat.

Het voorbeeld is natuurlijk erg simpel, maar het is eenvoudig uit te breiden tot meer realistische situaties. Vóór de inpassing beschikken we over N schattingen van N doelvariabelen, die ieder bestaan uit een brongegeven, een schatting van de vertekening in dat gegeven, en een schatting van de variantie van de samengestelde storingsterm bij deze schatting. De set van doelvariabelen moet aan $M < N$ balansrestricties voldoen. Dat betekent dat er $N - M$ doelparameters vrij geschat kunnen worden. Als de balansrestricties lineair zijn, resulteert substitutie van de balansrestricties in een gewogen lineair kleinste kwadratenprobleem, dat ook voor grote stelsels eenvoudig kan worden opgelost. Standaardfouten van onze uiteindelijke schattingen krijgen we daarbij cadeau.

In de hier geschetste opzet moet eerst per bron de vertekening geschat worden voorafgaand aan de confrontatie met andere bronnen. Als we de vertekening schatten door vast te kijken naar de andere bronnen, klopt het verhaal formeel statistisch niet meer. We moeten ons in dit stadium beperken tot wat we van die bron zelf (menen te) weten. Als het een basisstatistiek betreft, dan is idealiter een voldoende informatieve kwaliteitsrapportage beschikbaar (zoals onder meer Eurostat die tegenwoordig verlangt). Deze rapportage geeft niet alleen inzicht in de omvang van de steekproeffout, maar documenteert evenzeer alle beschikbare evidentie over de diverse vormen van niet-steekproeffouten. Bij administratieve bronnen is een dergelijke kwaliteitsrapportage niet aanwezig; idealiter vindt er dan flankerend steekproefsgewijs onderzoek plaats om de kwaliteit van deze bron in kaart te brengen. Dat kan door middel van eigen veldwerk, of door individuele koppeling aan (betrouwbare) gegevens die eerder door eigen waarneming zijn verkregen.

Bij het opstellen van onze schatting van de (verdeling van de) vertekening mogen we naast beschikbare kwaliteitsrapportages wel gebruik maken van onze kennis uit het verleden over de kwaliteit van de desbetreffende bron. Mits we toestaan dat deze 'vooroordelen' door nieuwe informatie gewijzigd kunnen worden. En dat leerproces ook zorgvuldig documenteren zodat het voor derden controleerbaar is; in technische zin is immers de aanname dat de vertekening zuiver geschat kan worden essentieel en moet daarom beargumenteerd worden waarom dat zo is. Een formeel statistisch mechanisme om dit te implementeren zijn zogenaamde Bayesiaanse schattingsmethoden. Het voert te ver om die hier technisch uit te werken; zie Magnus e.a. (2000) en VN (2002, hoofdstuk 6D) voor methoden en voorbeelden. We volstaan met de opmerking dat in dat geval zowel de vertekening als de doelparameters simultaan geschat worden bij de uiteindelijke integratie.

Gezien de omvang en complexiteit van de bestaande integratiestelsels, is het overigens twijfelachtig of deze technieken in formele zin hanteerbaar zullen zijn. We kunnen dan het proces ook stapsgewijs opzetten. Dat veronderstelt dat er een hiërarchische gelaagdheid is in het totale schattingsproces waarbij we laag voor laag kunnen afwerken. Per laag vinden partiële integraties plaats voor deelaspecten waarbij de bronnen die over het desbetreffende aspect informatie geven worden samengevoegd. Een dergelijke stapsgewijze opzet heeft het grote voordeel dat deze veel beter beheersbaar is. Uitkomst van elke laag is een schatting met een expliciete betrouwbaarheidsindicatie (vertekening, standaardfout) die vervolgens in de naasthogere laag kan worden ingezet. Van belang is dat door het hele proces heen niet alleen de puntschattingen gemanipuleerd worden, maar simultaan daarmee ook de betrouwbaarheid die daaraan lokaal (dat wil zeggen, gegeven de tot op dat moment verwerkte bronnen) moet worden verbonden.

De micro-integratie, door Henk van Tuinen integratie van onderop genoemd, moet in dit proces ook een plaats krijgen. Bij de meest simpele vorm van disseminatie van statistieken worden de gegevens en hun omschrijvingen als een grote database, een datawarehouse, gepubliceerd en stelt de gebruiker zijn eigen tabellen samen. Als deze gegevens consistent en vergelijkbaar zijn en hun begrippen geharmoniseerd, dan heeft de gebruiker daarmee zijn eigen samenhangende systeem. Door het CBS zullen in de komende jaren stappen in deze richting gezet worden met de ontwikkeling van

het Economisch statistisch bestand (ESB) voor de bedrijfsgegevens, het Sociaal statistisch bestand (SSB) voor de persoonsgegevens en de NR voor de overkoepelende gegevens. Samenhang op tabel-niveau, dus op meso-niveau, is haalbaar door zodanige wegingsfactoren te kiezen dat de tabel consistent is met bepaalde macrogegevens en met eerder gepubliceerde tabellen (Renssen en Nieuwenbroek, 1997; Kroese, Renssen en Trijssenaar, 2000; Renssen, Kroese en Willeboordse, 2001). Op deze manier kan een complete meso-macro-link ontwikkeld worden, waarbij uitgaande van zeer fijne basisindelingen, op meerdere manieren tabellen met verschillende indelingen gemaakt kunnen worden.

5. Conclusie

De achilleshiel van de integratiemethodologie ligt in de empirische verankering van de schatting van de (verdeling van de) vertekening per bron. Het volstaat niet hierbij te verwijzen naar de uitkomst van de confrontatie met andere, eveneens vertekende bronnen. Als meerdere bronnen van onbekende kwaliteit hetzelfde beeld vertonen biedt dat geen garantie dat dit beeld juist is. Er is zelfs geen garantie dat de vertekening kleiner wordt bij het combineren van dergelijke bronnen. Baron von Münchhausen kon zich aan zijn haren uit het moeras trekken, zulks is integratoren helaas niet gegeven. Zorgvuldig op empirie gebaseerd onderzoek naar de aard en omvang van de vertekening in de gebruikte bronnen is het enige wat er op zit. Het transparant en expliciet statistisch modelleren van de empirische relatie tussen de beschikbare bronnen en de doelvariabelen van de integratie, gebruikmakend van formele modellerings- en schattingstechnieken uit de wiskundige statistiek, biedt de enige grondslag voor vertrouwen in de uitkomsten van het integratieproces. Alleen zo kunnen wij de nauwkeurigheid van de schattingen bewaken en documenteren.

Literatuur

- ABS (Australian Bureau of Statistics), 2000. Australian National Accounts: concepts, sources and methods. Australian Bureau of Statistics, Canberra.
- Akritidis L., 2002. Accuracy assessment of National Accounts statistics. *Economic Trends* 589, 38–53.
- Bloem A.M. en S. Khawaja, 2001. A framework for assessing the quality of national accounts estimates. Internationaal Monetair Fonds, Washington D.C.
- CBS (Centraal Bureau voor de Statistiek), 1953. Nationale jaarrekeningen 1948-1950. De Haan, Utrecht.
- CBS (Centraal Bureau voor de Statistiek), 1992. Arbeidsrekeningen 1988. SDU, Den Haag.
- Cochran W.G., 1977. Sampling techniques. 3e druk. Wiley, New York.
- Cochran W.G. en G.M. Cox, 1950. Experimental designs. 2e druk, 1957. Wiley, New York.
- Edwards A.W.F., 1972, Likelihood. 2e druk, 1992. Johns Hopkins University Press, Baltimore.
- FCSM (Federal Committee on Statistical Methodology), 2001. Measuring and reporting sources of error in surveys. Statistical Policy Working Papers, no. 31. Office of Management and Budget, Washington D.C.

- Fisher R.A., 1925. *Statistical methods for research workers*. 14e druk, 1970, Hafner, New York.
- Fisher R.A., 1956. *Statistical methods and scientific inference*. 3e druk, 1973, Hafner, New York.
- Hacking I., 1965. *The logic of statistical inference*. Cambridge University Press, Cambridge.
- Kazemier B. en R. van Rooijen, 2002. *Assessment of the reliability of the Dutch provisional National Accounts*. Centraal Bureau voor de Statistiek, Voorburg / Heerlen.
- Keuning S.J. en K. Zeelenberg, 2002. *Samenhang in de statistiek*. Paper gepresenteerd op het Symposium Een informatiesysteem voor de Nederlandse economie, Den Haag, 9 september 2002. Centraal Bureau voor de Statistiek, Voorburg / Heerlen.
- Kroese, A.H. en R.H. Renssen R.H., 2000. Weighting or imputation: constructing a consistent set of estimates based on data from different sources. *Netherlands Official Statistics* 15, 23–31.
- Magnus J.R., J.W. van Tongeren en A.F. de Vos, 2000. National accounts estimation using indicator analysis. *Review of Income and Wealth* 46, 329–50.
- OESO / IMF / ILO / CIS STAT, 2002. *Measuring the non-observed economy: a handbook*. Organisatie voor Economische Samenwerking en Ontwikkeling, Parijs.
- ONS (Office for National Statistics), 2000. *UK ESA95 Gross National Income inventory of methods*. Office for National Statistics, Londen.
- Ploeg F. van der, 1982. Reliability and the adjustment of sequences of large economic accounting matrices. *Journal of the Royal Statistical Society Series A* 145, 169–194.
- Renssen R.H., A. Camstra, C. Huegen, W.J.G. Hacking en R.D. Huigen, 1998. A model for evaluating the quality of business surveys. Centraal Bureau voor de Statistiek, Voorburg / Heerlen.
- Renssen R.H. en N.J. Nieuwenbroek, 1997. Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association* 92, 368–374.
- Renssen R.H., A.H. Kroese en A.J. Willeboordse, 2001. *Aligning estimates by repeated weighting*. Centraal Bureau voor de Statistiek, Voorburg / Heerlen.
- Särndal C.-E., B. Swensson en J. Wretman, 1992. *Model-assisted survey sampling*. Springer, New York.
- SNA, 1993. *System of National Accounts*. EU / IMF / OESO / VN / Wereldbank, Brussel / Washington D.C. / Parijs / New York.
- Stigler S.M., 1986. *The history of statistics: the measurement of uncertainty before 1900*. Harvard University Press, Cambridge (Mass.).
- Stone R., D.G. Champernowne en J.E. Meade, 1942. The precision of National Accounts estimates. *Review of Economic Studies* 9, 111–25.
- Tuinen, H.K. van, 1978. Integratie van economische statistieken. In: H.P.H. Brouwer e.a. (red.), *Denken en meten: statistische opstellen*. SDU, Den Haag, blz. 18–29.
- VN (Verenigde Naties), 2002. *Handbook of macro accounts in policy analysis*. Verenigde Naties, New York.
- Wroe D., P. Kenny, U. Rizki en I. Weerakkody, 1999. *Reliability and quality indicators for National Accounts aggregates*. Office for National Statistics, Londen.
- York R. en P. Atkinson, 1997. The reliability of quarterly national accounts in seven major countries: a user's perspective. *Economics Department Working Papers* no. 171. Organisatie voor Economische Samenwerking en Ontwikkeling, Parijs.

Noten in de tekst

- ¹⁾ Het System of National Accounts zegt hierover: *"The design and structure of the System draws heavily on economic theory and principles as well as business accounting practices. ...When business accounting practices conflict with economic principles, priority is given to the latter, as the System is designed primarily for purposes of economic analysis and policy-making."* (SNA, 1993, blz. 11).
- ²⁾ Een soortgelijke indeling van integratiestappen is te vinden in de *National Accounts Process Table* van Eurostat (zie OESO, 2002, blz. 65-6). De beschrijving van de verschillende stappen is gebaseerd op de nog in ontwikkeling zijnde Nederlandse methodebeschrijving van het NR-proces ten behoeve van het BNP-Comité.

Peter Kooiman, Hoofd hoofdafdeling Modellen, arbeid en inkomen, Centraal Planbureau

Kees Zeelenberg, Hoofd sector Onderzoek en ontwikkeling, divisie MSP

Wouter J. Keller, Directeur M&I/Argitek, hoogleraar Vrije Universiteit
