

Economen in de fout met negeren meetfouten

Meetfouten in de economie zijn omvangrijk maar worden zelden of nooit transparant gerapporteerd. Wat voor impact heeft het negeren van deze meetfouten op economische analyses en beleid?

IN HET KORT

- De flash-raming van de Nederlandse bbp-groei zit er gemiddeld 0,5 procent naast. Bij een gemiddelde groei van 1,5 procent is dat veel.
- Financieel-economisch beleid dat geen rekening houdt met meetfouten kan te stevig of juist niet stevig genoeg reageren.
- Voor de geloofwaardigheid en om beleidsfouten te voorkomen, is het nodig transparanter te worden over meetfouten.

PETER VAN BERGEIJK
Emeritus hoogleraar aan de Erasmus Universiteit Rotterdam en directeur van Verbeeldinkt & olie

Om geloofwaardigheid te behouden en onhoudbare interpretaties en beleidsfouten te voorkomen, is het nodig transparanter te worden over meetfouten. Wie de bijstellingen binnenhaalt van de ramingen van bruto binnenlands product (bbp) door het Centraal Bureau voor de Statistiek (CBS) en ermee aan het rekenen slaat, ontdekt een opmerkelijke toename van de nauwkeurigheid van het statistische materiaal in het jaar 2016. Tot en met 2015 rapporteert het CBS de groei van het bbp op één decimaal nauwkeurig; dan is er plotsklaps een sprong in het onderliggende materiaal naar veertien (!) decimalen. Deze decimalen-tsunami is een mooi voorbeeld van het verschijnsel van ‘ongeloofwaardige nauwkeurigheid’ (Manski, 2011): de onder economen wijdverbreide gewoonte om hele reeksen decimalen te rapporteren. Deze pseudonauwkeurigheid is onzinnig en au fond amusant, maar tevens een symptoom van het onderliggende probleem dat economen nauwelijks aandacht besteden aan de onnauwkeurigheid van hun gegevens. Wie de meetfouten van het waarnemen van ons kenobject negeert, ontnemt zichzelf en anderen het zicht op de economische werkelijkheid.

Een kanttekening vooraf is op zijn plaats. De focus in dit artikel op het CBS is gekozen voor het Nederlandse lezerspubliek van *ESB*, maar het moge duidelijk zijn dat het probleem van de onnauwkeurigheid van economisch-statistisch materiaal geen Hollandse Ziekte is (Van Bergeijk, 2024). Nederland is eerder een bijzonder geval in de zin dat ons land een uitstekend statistiekbureau met hooggekwalificeerde medewerkers heeft met redelijk wat financiële slagkracht en een enorme ervaring. Dat is extra ontvullend:

in veel landen is het statistische apparaat minder goed ontwikkeld en zullen foutmarges navenant groter zijn.

Nederland als voorbeeld

Bij de introductie van de nationale rekeningen in Nederland was het CBS voor het eerst en laatst transparant over de onnauwkeurigheid van de bbp-raming en gaf het bureau per component een kwantitatieve indicatie (CBS, 1953). De meest nauwkeurige waarnemingen (bijvoorbeeld voor de indirecte belastingen of de loonsom) hadden een gerapporteerde onnauwkeurigheid van twee tot vijf procent; het minst nauwkeurig werden de bedrijfsreserves waargenomen (foutmarge meer dan twintig procent; er is overigens eigenlijk geen reden waarom de nauwkeurigheid voor deze categorie in de loop der tijd zou zijn verbeterd).

Het is bij zulke onnauwkeurigheden overduidelijk dat het rapporteren van economische groei op één decimaal nauwkeurig wetenschappelijk gezien een dubieuze praktijk is. Immers, een ‘ware’ groeivoet van bijvoorbeeld drie procent kan bij een meetfout van twee procent van de begin- en eindwaarden worden waargenomen als ergens tussen minus één procent en plus zeven procent.

Omdat statistiekproducenten in het algemeen niet transparant zijn over hun onnauwkeurigheden, werken we met een indirecte waarneming, namelijk de vergelijking van wijzigingen die in de loop der tijd optreden in de ramingen voor een specifiek jaar. Omdat iedere raming een meetfout kent, levert deze vergelijking logischerwijze een conservatieve, minimale indicatie van de onnauwkeurigheid op.

Het idee dat de nauwkeurigheid van bbp-ramingen in de loop der tijd is toegenomen, is wijdverbreid, maar helaas een illusie. Over de periode 2001–2022 belooft het gemiddelde absolute verschil tussen de eerste ‘flash’-raming van het bbp (45 dagen na het afsluiten van een jaar) en de definitieve raming (twee jaar later) een half procentpunt (figuur 1). Bij een gemiddelde groeivoet van 1,5 procent per jaar is de onnauwkeurigheid dus een derde deel. Dit is een opmerkelijke stijging ten opzichte van eerdere bepalingen van de onnauwkeurigheid van de CBS-ramingen: Geelhoed en Van Sinderen (1993) rapporteerden dat de gemiddelde meetfout van bbp-groei ruwweg een kwart was van de gemiddelde groeivoet in de jaren 1971–1992; voor 1993–2005 was dit ruim een vijfde (Van Bergeijk, 2010).

De verschillen in figuur 1 lijken klein en het conjunctuurbeeld wordt visueel niet al te veel geweld aangedaan (in 2003 heeft de flash het verkeerde teken en in 2013 wordt de conjunctuuromslag te laat waargenomen). Een gedachtenexperiment toont dat zulke kleine verschillen er wel degelijk toe doen. Als we twintig jaar lang de flashraming als



waar en vaststaand aannemen en als finale raming gebruiken, leiden deze kleine jaarlijkse verschillen uiteindelijk tot een verschil met de definitieveraming van tien procentpunt (figuur 2). Daarbij moet worden aangetekend dat het definitieve cijfer simpelweg het cijfer is dat het CBS niet meer herziet. Dat wordt even goed onnauwkeurig waargenomen.

Het bbp-niveau is medebepalend voor belangrijke indicatoren zoals de schuldratio en het overheidstekort waarvoor (inter)nationale normen zijn afgesproken. Bijsturing van de economie vereist tijd, en financieel-economisch beleid dat wordt gebaseerd op de vroege flash-raming kan door vertekeningen op langere termijn op het verkeerde been worden gezet.

Waarom de nauwkeurigheid niet verbetert

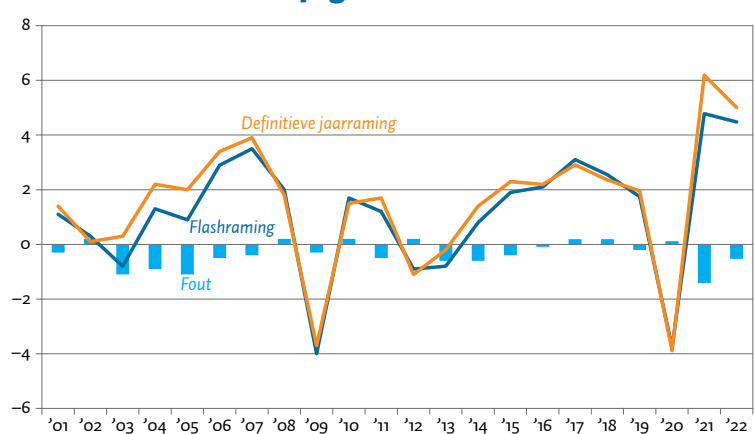
Het lijkt verwonderlijk dat de nauwkeurigheid van de ramingen niet verbetert, ondanks de door de digitale revolutie enorm toegenomen reken capaciteit en nieuwe registratiemogelijkheden, maar het is een algemeen verschijnsel waarvoor plausibele verklaringen bestaan (Van Bergeijk, 2024).

In de eerste plaats is de paradox van Simpson relevant: de ontwikkeling van het totaal kan door verschuivende aandelen van de componenten tegengesteld zijn aan een ondubbelzinnige ontwikkeling van de samenstellende delen. Voor ieder van de onderdelen kan de waarneming in de tijd dan verbeteren terwijl het aggregaat verslechtert. Aangezien het aandeel van moeilijk waarneembare activiteiten toeneemt, verbetert de nauwkeurigheid op macroniveau niet (Griliches, 1994). De dynamiek van markteconomieën, het toenemende belang van diensten, innovatie en (de)mondialisering leveren voortdurend uitdagingen voor het statistische apparaat op.

Een tweede verklaring is de aandacht die gegevensproducenten zijn gaan geven aan het zo snel en vroeg mogelijk aanleveren van statistieken, ook om zo de nieuws waarde

Flash-raming van bbp-groei vergeleken met de definitieve bbp-groei

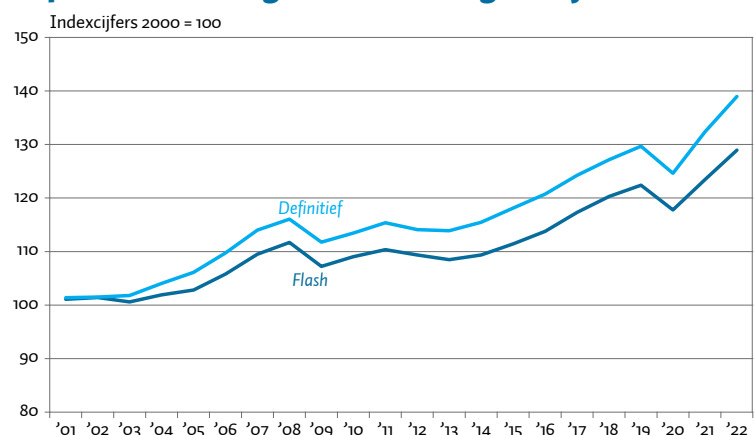
FIGUUR 1



Data: CBS | ESB

Bbp uit flashraming en definitieve groeicijfers

FIGUUR 2



Data: CBS | ESB

van hun cijfers te vergroten en daarmee het maatschappelijk draagvlak te versterken. Marketing en versnelde productie leggen beslag op middelen en mensen; bij budgetten die onder druk staan, is het logisch dat dit leidt tot minder ruimte voor andere karakteristieken, zoals de nauwkeurigheid (Fixler, 2007). Hier spelen twee zaken: ten eerste het ontbreken van externe monitoring van de nauwkeurigheid van statistieken omdat meetfouten niet worden gerapporteerd, en ten tweede de tendens om de beschrijving van de kwaliteit van statistieken in toenemende mate te richten op procesbeschrijving (Kooiman et al., 2003). De popularisering van de statistiek doet onverantwoorde concessies aan de wetenschappelijke kwaliteit en leidt tot verwarstatistiek: persberichten die aan volstrekt insignificante veranderingen achter de komma belang toekennen, en krimp en groei binnen de foutmarge ten onrechte als een belangrijk nieuwswaardig item aan de persoon brengen.

Het is natuurlijk ook mogelijk dat de onnauwkeurigheid van statistieken helemaal geen probleem wordt gevonden. De econometrie heeft technieken ontwikkeld om te corrigeren voor zogenoemde klassieke meetfouten en dan doen de fouten in het bronmateriaal, zo wordt gesteld, er uiteindelijk niet toe. Er zijn echter steeds meer indicaties dat meetfouten zich niet 'klassiek' gedragen, bijvoorbeeld doordat ze groter zijn in een crisis. Dan leiden de foutcorrecties juist tot extra vertekening (zie ook figuur 1 en 2).

Daarbij komt dat beleidmakers vooral geïnteresseerd zijn in de meest recente gegevens. Dat deze cijfers ook nog jaren later worden herzien, mag zich in weinig beleidsmatige aandacht verheugen. Indien het gedrag van beleidmakers metterdaad door de vroegst beschikbare cijfers wordt bepaald, dan is het voor de analyse van hun gedrag minder relevant om te kijken naar de nauwkeurigere latere ramingen.

Ten slotte lijkt onder statistiekproducenten het idee te leven dat het beter is om meetfouten alleen te rapporteren als deze ook nauwkeurig kunnen worden vastgesteld. Er is ook een zekere huiver omdat transparantie ten koste zou kunnen gaan van het vertrouwen van het publiek in de statistieken. De paradox is dat door het verhullen van de onnauwkeurigheid onvoldoende middelen worden aangewend ter verbetering van de statistieken, terwijl het risico op reputatieschade verder toeneemt als de nauwkeurigheid afkalt.

Relevantie van meetfouten

De meest gemaakte tegenwerping tegen pleidooien om meetfouten transparant te rapporteren is dat meetfouten in het bronmateriaal verbleken in het licht van de modelonzekerheid en de in modellen gedane aannames. Dat is een terecht punt, natuurlijk. Ten tijde van de coronapandemie liepen de ramingen van het CBS en het Centraal Planbureau (CPB) voor de groei van de overheidsbestedingen bijvoorbeeld uiteen van 1,1 tot 1,5 procent (voor de overheidsinvesteringen van 1,9 tot 4,5 procent). En hoewel dat een ongekend groot verschil was, bleek de modelonzekerheid vele malen groter. Doorrekening in het CPB-model Saffier levert bij deze impuls een groeivoet voor het bbp op van 0,7 tot 1,2 procent; met het model Delphi van De Nederlandsche Bank is dat 0,2 tot 0,3 procent. Een duidelijk voorbeeld dat de applicatiefout de meetfout in het

bronmateriaal vele malen kan overtreffen. Het is geen op zichzelf staand voorbeeld; grote modelonzekerheid is bijvoorbeeld ook aan de orde gesteld rond de doorrekening van de verkiezingsprogramma's (Haffner en Van Bergeijk, 1994) en de (in)effectiviteit van belastingverschuivingen in een veelvoud aan beleidsmatig gebruikte modellen (Van Bergeijk, 1994).

Echter, de betekenisvolle omvang van applicatiefouten betekent niet dat de onnauwkeurigheid van het bronmateriaal er niet toe doet. In de eerste plaats hebben fouten de neiging zich te verspreiden en ieder onderdeel van de uiteindelijke onnauwkeurigheid doet er dus toe.

Ten tweede is met name de beschrijvende statistiek in de beleidsvoorbereiding belangrijk, ook omdat recente gegevens de beeldvorming bepalen. Hier blijken meetfouten veel groter te zijn dan onderkend, en liggen beleidsfouten op de loer zoals blijkt uit de volgende drie voorbeelden.

Duurzame ontwikkelingsdoelen

Omvangrijke datagestuurde projecten zoals de duurzame ontwikkelingsdoelen (SDG's) van de Verenigde Naties worden als een grote stap voorwaarts gezien, maar omdat er bij de selectie van gekwantificeerde doelstellingen geen enkele aandacht is besteed aan de (on)nauwkeurigheid van de doelvariabelen, is teleurstelling als het ware ingebakken. Een goed voorbeeld is de doelstelling om de mondiale armoede terug te dringen. Prydz et al. (2022) vergeleken het inkomen per hoofd op basis van 2.000 nationale huishoudsurveys (166 landen) in de periode 1965–2019 met de gegevens op basis van nationale rekeningen, en vinden dat de inkomens volgens de nationale rekeningen vijftig procent hoger zijn. Ze berekenen voor het jaar 2011 dat de armoededoelstelling van minder dan drie procent wereldwijde armen op basis van de nationale rekeningen werd bereikt, maar dat dat aandeel dubbel zo hoog was als naar de bevragingen van de huishoudens werd gekeken. Wie de onnauwkeurigheid niet rapporteert, dreigt op haar lauweren te gaan rusten.

Inflatie

Zelfs topeconomen, waaronder Blanchard (voormalig IMF-hoofdeconoom), DeLong (voormalig staatssecretaris Treasury in de VS), en de Nobelprijswinnaars Sargent en Krugman houden in hun beleidsadvies geen rekening met meetfouten (Bolhuis et al., 2022). Op basis van de zeer scherpe reactie van inflatie op het aantrekken van de monetaire teugels door FED-voorzitter Volcker in de jaren tachtig adviseerden zij dat er maar een lichte monetaire verkrapping nodig zou zijn om de inflatie door de pandemie en door het uitbreken van de oorlog van Rusland tegen Oekraïne in het gareel te krijgen. Daarbij zagen zij over het hoofd dat de inflatie in de jaren tachtig anders gemeten werd, waardoor de reactie van inflatie op de monetaire verkrapping groter leek. Wie een correcte vergelijking met de jaren tachtig wil maken houdt daar rekening mee en concludeert dat de recente stevigere monetaire verkrapping noodzakelijk was.

Dankzij het goede werk van Suyker (2024) weten we dat we de inflatie ook in Nederland anders zullen moeten

beoordelen dan de vroegere cijfers suggereerden (ze steeg minder hard en bleef langer hoog, zie ook Chowdhury en Dixon (2024) voor het lockdownjaar 2020). De inflatiediscussie illustreert en bevestigt mijn punten dat onbegrip van onnauwkeurige metingen endemisch is in de professie. Beleidsadviezen moeten én genuanceerd én geamendeerd worden, indien met de onnauwkeurigheden rekening wordt gehouden.

Wereldproductie

Het derde voorbeeld is de ontwikkeling van de wereldproductie rond de grote wereldwijde crises, zoals de Volckerrecessie uit het begin van de jaren tachtig (die volgens recente data geen krimp was), de Grote Recessie van de bankencrisis (waarvan nog steeds onduidelijk is of er metterdaad krimp is opgetreden) en de Grote Lockdown waarvan de grootte van de krimp sterk is overschat (in de Verenigde Staten bijvoorbeeld met een derde). In al deze gevallen zijn de eerste ramingen stelselmatig te negatief geweest en de beleidsreactie navenant te sterk, zie ook Van Bergeijk (2024).

Conclusie en implicaties

Meetfouten doen er dus toe. Zowel voor het beleid als voor de economische analyse van actuele gebeurtenissen, maar ook voor analyses die betrekking hebben op lange perioden waarin meetfouten fluctueren. Als we die meetfouten transparant rapporteren, helpt dat om genuanceerder te oordelen over de context en om geen conclusies te trekken die niet door de gegevens gedragen kunnen worden.

Het is een illusie te verwachten dat we nauwkeuriger zullen gaan meten. De meetfouten die Morgenstern in 1950 in zijn *On the accuracy of economic observations* aan de orde stelde, blijken in recent onderzoek van eenzelfde orde-grootte te zijn gebleven (Manski, 2015; Linsi en Mügge, 2019; Van Bergeijk, 2024). In de jaren zestig en zeventig werd de hoop gevestigd op de mogelijkheden van digitale datacollectie en de enorm toenemende reken capaciteit van computers, maar een noemenswaardige verandering trad niet op.

Tegenwoordig is de hoop gevestigd op kunstmatige (AI) intelligentie en big data, maar deceptie ligt ook hier op de loer. Het trainingsmateriaal voor AI-modellen is immers onvolkomen, omdat meetfouten momenteel niet worden gerapporteerd. AI-modellen zullen dus de neiging hebben om simpelweg de huidige slechte praktijken te herhalen en ook nog te gaan hallucineren wanneer gegevens ontbreken.

Bovendien speelt de paradox van de big data op (Meng, 2016) dat het vergroten van de steekproefomvang weliswaar meer nauwkeurigheid suggereert omdat de grote hoeveelheid waarnemingen van nieuwe bronnen de statistische betrouwbaarheidsintervallen verkleint, maar tegelijkertijd de *bias* van onvolkomen steekproeven vergroot. Dan geldt meestal: 'hoe meer gegevens hoe zekerder we onszelf voor de gek houden'.

De enige manier om onhoudbare interpretaties en beleidsfouten te voorkomen en op langere termijn geloofwaardigheid te behouden – en daarmee draagkracht voor het beleid – is om transparant te worden over meetfouten. Dit is een inzicht dat overigens breed gedeeld wordt in

andere wetenschappen. Er is geen reden waarom de economische analyse hier een uitzonderingspositie zou moeten innemen.

Het is hoopgevend dat er beleidsmatige interesse aan het ontstaan is om in navolging van de grote studies over de opwarming van het klimaat ook te bestuderen wat de mogelijkheden zijn om het publiek beter te informeren over de onzekerheid en onnauwkeurigheid van meetresultaten (Van der Bles et al., 2019; De Jonge 2020) en dat transparantie daarover geen negatieve invloed heeft op het vertrouwen van professionele gebruikers en het publiek (Galvão en Mitchell, 2023). Het blijft echter ontnuchterend dat economen in het algemeen in de fout gaan.

Literatuur

- Bergeijk, P.A.G. van (1994) Op zoek naar een gratis lunch. *ESB*, 79(3984), 1013–1017.
- Bergeijk, P. van (2010) Meten en nog steeds niks weten. *ESB*, 95(4590), 463.
- Bergeijk, P.A.G. van (2024) *On the inaccuracies of economic observations: Why and how we could do better*. Cheltenham: Edward Elgar.
- Bles, A.M. van der, S. van der Linden, A.L.J. Freeman et al. (2019) Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, 6(5), 181870.
- Bolhuis, M.A., J.N.L. Cramer en L.H. Summers (2022) Comparing past and present inflation. *Review of Finance*, 26(5), 1073–1100.
- CBS (1953) *Nationale jaarrekeningen 1948–1950*. CBS. Te vinden op historisch.cbs.nl.
- Chowdhury, A. en H. Dixon (2024) Measuring inflation during the pandemic with the benefit of hindsight. *Open Economies Review*, (2024), 1–26.
- Fixler, D. (2007) Timeliness and accuracy. In: M. Boumans (red.), *Measurement in economics: A handbook*. Amsterdam: Elsevier, p. 413–427.
- Galvão, A.B. en J. Mitchell (2023) Communicating data uncertainty: Multiwave experimental evidence for UK GDP. *Journal of Money, Credit and Banking*, 56(1), 81–114.
- Geelhoed, L.A. en J. van Sinderen (1993) National accounts, economic-policy standards and economic policy: Some observations. In: W.F.M. de Vries, G.P. den Bakker, M.B.G. Gircour et al. (red.), *The value added of national accounting: Commemorating 50 years of national accounts in the Netherlands*. Voorburg/Heerlen: CBS, p. 167–190. Te vinden op historisch.cbs.nl.
- Griliches, Z. (1994) Productivity, R&D, and the data constraint. *The American Economic Review*, 84(1), 1–23.
- Haffner, R.C.G. en P.A.G. van Bergeijk (1994) The economic consequences of Dutch politics. *De Economist*, 142(4), 497–505.
- Jonge, E. de (2020) *Communicating uncertainties in official statistics: A review of communication methods*. Eurostat Statistical Working Paper. Te vinden op ec.europa.eu.
- Kooiman P., K. Zeelenberg en W.J. Keller (2003) De nauwkeurigheid van geïntegreerde gegevens. In: S.A. Algera, P. van der Laan en J. Nobel (red.), *Gedacht en gemeten*. Voorburg/Heerlen: CBS, p. 111–120.
- Linsi, L. en D.K. Mügge (2019) Globalization and the growing defects of international economic statistics. *Review of International Political Economy*, 26(3), 361–383.
- Manski, C.F. (2011) Policy analysis with incredible certitude. *The Economic Journal*, 121(554), F261–F289.
- Manski, C.F. (2015) Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. *Journal of Economic Literature*, 53(3), 631–653.
- Meng, X.-L. (2018) Statistical paradises and paradoxes in big data: Law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics*, 12(2), 685–726.
- Morgenstern, O. (1950) *On the accuracy of economic observations*. Princeton: Princeton University Press.
- Prydz, E.B., D. Jolliffe en U. Serajuddin (2022) Disparities in assessments of living standards using national accounts and household surveys. *Review of Income and Wealth*, 68(52), S385–S420.
- Suyker, W. (2024) *De correcte Nederlandse inflatiecijfers in 2021–2024*. Statistiek op esb.nu, 12 juli.