

Datavisualisatie in de MEV onvoldoende

In de economische wetenschap is datavisualisatie in opkomst. Gegevens moeten correct en begrijpelijk worden weergegeven. De figuren in de MEV blijken onvoldoende te scoren qua deugdelijkheid, leesbaarheid en begrijpelijkheid. Daardoor kunnen analyses verkeerd worden begrepen en verkeerde beleidsconclusies worden getrokken.

FRANK KALSHOVEN
Directeur bij de
Argumentenfabriek

PETER VAN BERGEIJK
Hoogleraar aan
het Institute of
Social Studies aan de
Erasmus Universiteit Rotterdam
en lithograaf bij
Verbeeldinkt&olie

Schwabisch (2014) publiceerde onlangs in de *Journal of Economic Perspectives* zijn 'An economist's guide to visualizing data'. Zijn hoofdstelling: als economen hun kennis willen verspreiden, moeten ze tijd investeren in overtuigende, effectieve grafieken. Schwabisch hanteert drie vuistregels. Eén: laat de gegevens zien. Twee: dring de 'ruis' in grafieken terug. Drie: integreer tekst en grafiek.

De Macro-Economische Verkenningen (MEV) is het 'topstuk' onder de beleidsdocumenten die jaarlijks in Nederland gepubliceerd worden. Het heeft een grote impact op politiek, ambtenaren en journalisten. Daarom kozen we de MEV 2015 voor een vingeroefening met datavisualisatie. Voldoen de figuren in de MEV aan de eisen die hieraan mogen en moeten worden gesteld?

WESTEINDER-INDEX

Er bestaan al wel richtlijnen voor illustraties in wetenschappelijke publicaties, zoals die van de American Psychological Association (APA), maar een methode om grafieken onderling te vergelijken, is nog niet eerder ontwikkeld. Om de kwaliteit van de figuren in de MEV te beoordelen, leggen we ze daarom langs een zelf geconstrueerde meetlat. Die meetlat noemen we

naar de plaats waar hij is ontwikkeld, de Westeinderplassen.

Ons vertrekpunt zijn de APA-vereisten. Anders dan de APA, vinden we het nummeren van figuren, verwijzen in de tekst, correcte spelling en een leesbaar letterkorp geen punten waard, want tekstverzorging is belangrijk maar nogal basaal. We toetsen ook niet of de figuur *noodzakelijk* is voor het betoog. We beoordelen de figuren als zelfstandige argumenten. Figuren uit de MEV worden vaak geplaatst in de krant en beleidsstukken waarbij de MEV-context ontbreekt, en daarom moeten ze dus zonder context leesbaar zijn.

De Westeinder-index bestaat uit twee hoofdonderdelen: 'deugdelijkheid' en 'leesbaarheid en begrijpelijkheid'. Vier APA-vereisten komen terug bij 'deugdelijkheid' (assen, eenheden, begrippen en afkortingen, bronvermelding). We voegen een beoordeling van het grafiekbijchrift toe: dubbelzinnige en verwarrende titels leiden tot verkeerde interpretaties. Onder 'leesbaarheid en begrijpelijkheid' nemen we het APA-criterium 'afwezigheid van ruis' mee en beoordelen daarnaast de boodschap en de relaties die tussen figuren worden gelegd. Onze keuze van beoordelingscriteria heeft invloed op de uitkomsten. Over die keuze is verschil van inzicht mogelijk waar het gaat over de leesbaarheid en begrijpelijkheid; voor het deugdelijkheidsaspect zien wij daarvoor minder ruimte.

We gebruiken een kwalitatieve benadering en scoorden alle onderdelen van de index onafhankelijk van elkaar. We gaven nul punten (slecht gedaan, ontbreekt) of 1 punt (voldoende, goed gedaan) op alle acht aspecten van de Westeinder-index, en bespraken daarna de afwijkende scores om tot een definitief oordeel te komen. Kwamen we er onderling niet uit, dan werd het minst strenge oordeel gevolgd. De score per figuur is zichtbaar in tabel 1.

De MEV bevat geen lijst van figuren; maar er staan er 24 in. Dit is een voldoende grote populatie om zinnige conclusies te kunnen trekken over de kwaliteit van datavisualisatie in de MEV en over de bruikbaarheid van de nieuwe index.

De MEV heeft figuren die goed scoren, maar ook ernstige tekortkomingen. De beste grafiek volgens de Westeinder-

De auteur heeft verklaard dit artikel alleen te publiceren in ESB en niet elders te publiceren in wat voor medium dan ook. Het is wel toegestaan om het artikel voor eigen gebruik en voor publicatie op een intranet van de werkgever van de auteur aan te wenden.

index is figuur 2.13. Deze figuur toont dat de Nederlandse reële loonvoet sneller groeit dan de arbeidsproductiviteit, waardoor onze arbeidskosten sneller stijgen dan elders: een ingewikkelde relatie die grafisch helder wordt geïllustreerd. Een duidelijke titel, ondubbelzinnige assen, variabelen en bronvermelding, geen overbodige informatie en de boodschap komt over. Zo moet het. Helaas houdt de MEV dit kwaliteitsniveau niet vol.

DEUGDELIJKHEID

De eerste set criteria betreft ‘deugdelijkheid’: het zijn de standaardeisen waaraan elke grafiek moet voldoen. Per criterium bespreken we hoe de MEV scoort en geven we voorbeelden.

Heldere titel

Laat de titel zien waar de figuur over gaat? Anders dan verwacht blijkt het maken van een grafiekopschrift geen sinecure. In ruim dertig procent van de gevallen gaat het mis: onduidelijkheid, dubbelzinnigheid of tegenstrijdigheid. We

zijn streng maar rechtvaardig. Het kader op pagina 13 presenteert ‘Volatiliteit DAX bij gewapende invallen (links) en reactie bbp eurozone en voor Nederland relevante wereldhandel op onzekerheidsimpuls in het eerste en tweede jaar daarna (rechts)’. Nul punten voor het bijschrift van een figuur die toont dat gewapende conflicten op de beurs wel opwinding veroorzaken, maar in de macro-economie slechts een kortetermijneffect hebben. ‘Kerngegevens collectieve sector 2001–2015’: saai, maar ok.

Afkortingen en begrippen

Bij bijna zeventig procent van de figuren zijn de afkortingen en begrippen voor ons als gebruikers duidelijk. Dat zou natuurlijk honderd procent moeten zijn. Het gaat al mis in figuur 1.1: ‘Relevante wereldhandel en uitvoer (links), en inflatie eurogebied blijft dalen (rechts)’. Wat bedoelt de MEV hier met relevante wereldhandel? Is dat de voor Nederland relevante wereldhandel (zo wordt het begrip vaak gebruikt) of de voor eurolanden relevante wereldhandel (logischer, aangezien het rechter paneel over het eurogebied gaat)? Grafieken moeten zonder nadere bestudering van de hoofdtekst begrijpelijk zijn: nul punten.

Eenheden

Wat staat er op de as? Een indexcijfer (en met welk basisjaar?), een percentage (en waarvan?) of is het een groeivoet? De MEV heeft er grote moeite mee: het gaat in ruim veertig procent van de gevallen mis. In figuur 2.8 bijvoorbeeld, over kredietvoorwaarden, staat op de as ‘% banken met wijziging’. Je zou denken, dat is dan iets tussen de 0 en 100 procent. Echter: de as loopt van –100 naar +100 (en wordt ook helemaal gebruikt). Het is raden wat hier wordt bedoeld. Nul punten.

Afgestemde assen

Door verkeerd gekozen assen kan de visuele boodschap het patroon in de gegevens vertekenen. Als er links een ontwikkeling is te zien voor 2008 tot 2014 ligt het in de rede dezelfde periode rechts te gebruiken. De MEV maakt van de assen een potje. In ruim zeventig procent van de gevallen zijn er gebreken. De lage score komt niet alleen door gebrekkige afstemming van de periode (horizontale as). Ook op de verticale assen is het een rommeltje. Als in twee panelen een ‘nul’ voorkomt, kiest de MEV er zelden voor die nul-assen ook op de zelfde hoogte te plaatsen. Als getallen zich in een ander bereik bevinden, kiest de MEV lukraak de reikwijdte van de as. In figuur 3.1 loopt de as voor collectieve uitgaven van 40 tot 50 procent en voor collectieve lasten van 30 tot 40 procent. Voor elk jaar in de grafiek zijn de uitgaven groter dan de lasten, maar wie niet heel scherp op de assen let, ziet het omgekeerde.

Bronvermelding

Om te kunnen controleren en te begrijpen wat er hoe is berekend, zijn bronvermeldingen noodzakelijk. Liefst specifiek. We scoorden niet de kwaliteit van de bronvermelding, maar louter diens aanwezigheid. Slecht 4 van de 24 afbeeldingen vermelden een bron. Bronvermelding is noodzakelijk want de Miljoenennota en de MEV rapporteren verschillende waarden voor tekort en schuldquote (De Kam, 2014). Het ontbreken klemt te meer omdat het CPB verschillende mo-

Score van de 24 MEV-figuren op de acht onderdelen van de Westeinder-index¹

TABEL 1

Figuur	Deugdelijkheid					Leesbaarheid en begrijpelijkheid			Score Westeinder-index
	Titel	Afkortingen en begrippen	Eenheden	Assen	Bron	Boodschap	Geen ruis	Relatie	
1.1									0
Kader (p.13)							V		1
1.2	V	V				V		V	4
1.3		V					V		2
1.4	V	V	V			V		V	5
1.5				V					1
2.1	V	V	V						3
2.2	V	V	V	V					4
Kader (p. 30)				V					1
2.3	V	V		V		V	V	V	6
2.4	V	V	V			V		V	5
Kader (p. 34)	V	V				V	V		4
2.5		V							1
2.6	V		V						2
2.7	V	V	V			V	V	V	6
2.8				V					1
2.9	V	V	V	V		V	V	V	7
2.10			V						1
2.11	V	V	V		V		V		5
Kader (p. 35)	V	V	V		V				4
2.12	V		V		V	V			4
2.13	V	V	V	V	V	V	V	V	8
3.1	V	V	V				V		4
3.2	V	V	V	V		V	V	n.v.t.	7

¹ De auteurs gaven nul punten (slecht gedaan) of één punt (goed gedaan) op alle acht aspecten van de Westeinder-index.

dellen heeft en het onduidelijk is welk model waar is ingezet. In het kader op pagina 13 is het belangrijk te weten wat voor soort model er is gebruikt. Is het een model à la SAFFIER of Worldscan of misschien nog iets anders? Het maakt nogal wat uit bij de interpretatie.

Totaalscore deugdelijkheid

De overall-score van de MEV-figuren op het eerste deel van de Westeinder-index is gemiddeld onvoldoende: 2,4 van de maximaal 5 punten. De spreiding is groot. Durven we de lat desondanks hoger te leggen? Ja, al was het maar om een 'nulmeting' te hebben, zodat we in de toekomst de voortgang kunnen meten en vergelijkingen kunnen maken met andere publicaties.

LEESBAARHEID EN BEGRIJPELIJKHEID

Eigenlijk komen we met deel twee van de Westeinder-index pas toe aan de wezenlijke functie van de grafiek: in een oogopslag laten zien wat de data de kijker leren, uiteraard op een integere manier.

Heldere boodschap

Van minder dan de helft van de figuren is de boodschap helder. Figuur 2.7 heeft een duidelijk verhaal 'Investeringsniveau trekt aan door toename van bezettingsgraad en producentenvertrouwen.' Het interessante van dit voorbeeld is dat deze figuur niet goed scoort in het eerste deel van de index. Een 'slechte grafiek' kan dus toch een heldere boodschap hebben. Het gaat vaak mis. Een opmerkelijke waarneming in de MEV is dat de beroepsbevolking krimpt bij voorzichtig economisch herstel (figuren 1.3, 2.5 en 2.11). In figuur 1.3 zien we de werkgelegenheid toenemen terwijl de beroepsbevolking krimpt. In figuur 2.11 zien we een rimpeling in de vacatures en een identieke krimpfiguur. Het rechterpaneel van figuur 2.5 komt nog het dichtst bij de door ons vermoede boodschap, en dat is ook waar het ons om gaat: drie verwarrende dubbelfiguren in plaats van de deelfiguur die de boodschap wel over kan brengen.

Geen ruis

Met 'ruis' in een grafiek bedoelen we de informatie die onnodig is om het patroon in de gegevens te laten zien en dat soms zelfs verhindert. Bijna zestig procent van de gevallen bevat te veel informatie om functioneel te kunnen zijn.

Figuur 1.2 'Bijdrage consumptie wordt positief (links) en inkomensgroei draagt consumptiegroei (rechts)' is een schoolvoorbeeld van ruis. Links verwacht de kijker de consumptieontwikkeling. En die staat er ook. Maar tevens de ontwikkelingen van alle andere finale bestedingscategorieën én het bbp. Allemaal ruis dus. Het rechterpaneel idem dito: een plaatje dat de inkomensgroei beschrijft, maar ook iets met vermogens, de eigenwoningwaarde en een 'rest/onverklaard'. Eindresultaat: de kijker blijft in verwarring achter. Zonde, want met zulke gegevens heb je als figurenbouwer goud in handen

Relatie tussen deelgrafieken

De MEV gebruikt figuren met twee en soms meer panelen. We beoordelen of de relatie tussen de deelgrafieken is geëxpliciteerd. Daar waar Criterium 1 al wordt gehaald als de titel verklaart wat we zien, moet bij Criterium 8 duidelijk worden

(in de titel) wat hun relatie is. De suggestie bij deelgrafieken is die van implicatie en mogelijk zelfs van causaliteit: 'omdat x (in luik 1), dus y (in luik 2)' of 'ondanks x toch y .' Dit gaat goed in onze favoriet, afbeelding 2.13. 'Reële loonvoet groeit harder dan arbeidsproductiviteit (links), waardoor arbeidskosten meer stijgen dan bij Europese concurrenten (rechts).' De relatie (oorzaak-gevolg) is expliciet gemaakt. Soms is het evident niet de bedoeling om een relatie te leggen; dan hebben we het criterium niet van toepassing verklaard.

Bij zeventig procent gaat het mis. Figuur 2.1 lijkt een onschuldig voorbeeld. 'Ongelijk herstel: bbp (links) en inflatie (rechts) in geavanceerde economieën.' Maar is dat wel zo, vraagt de kijker zich af. Of bedoelt de MEV eigenlijk dat de inflatie sneller oploopt naarmate het herstel sneller gaat? Dat zou kunnen, maar het laat zich dan weer uit het plaatje lastig lezen.

Totaalscore leesbaarheid en begrijpelijkheid

Op deel twee scoren de grafieken in de MEV gemiddeld 1,2 van de maximaal 3 punten. Hier is de spreiding nog groter: acht grafieken krijgen nul punten, maar het is hoopgevend dat drie grafieken de maximale score halen.

SLOT

Combineren we de twee subindices dan komt de gemiddelde score per grafiek uit op 3,6 van de maximaal 8 punten. Van de 192 grafiekaspecten (24 grafieken; de Westeinder-index heeft 8 elementen) is 1 niet van toepassing verklaard en 85 zijn op orde. De Westeinder-index indiceert daarmee 106 verbeterpunten.

Al met al lijkt de Westeinder-index geschikt om de kwaliteit van datavisualisatie te toetsen ongeacht de (tekstuele) context. Ofschoon we niet uitsluiten dat de Westeinder-index op onderdelen verbeterd kan worden, lijkt de huidige versie een nuttig instrument. Tot slot: dit is niet meer of minder dan een vingeroefening om een aanwijzing te krijgen over hoe bekwaam Nederlandse economen zijn in het visualiseren van data. Het antwoord valt ons tegen, want zelfs de MEV (een publicatie met een voorbeeldfunctie) scoort gemiddeld minder dan 50 procent. Dat kan beter. Het moet ook beter: beleidsmakers en journalisten hebben weinig tijd en zijn notoire grafiekensnellers. Verwarrende grafieken leiden tot verkeerde beslissingen.

LITERATUUR

- APA (2010) *Publication manual of the American Psychological Association*. Manual op www.apa.org.
- Centraal Planbureau (2014) *Macro Economische Verkenning 2015*. Den Haag: CPB.
- Kam, F. de (2014) Omissies in de Miljoenennota 2015. *ESB*, 99(4965), 584–586.
- Schwabish, J.A. (2014) An economist's guide to visualizing data. *Journal of Economic Perspectives*, 28(1), 209–233.