

De politieke relevantie van correlaties

J. de Leeuw

0 Inleiding

In hun in dit nummer van de Sociologische Gids gepubliceerde artikel passen Dronkers en De Jong (voortaan D&dJ) het model van Jencks (1972) toe op Nederlandse gegevens. Hun uitkomsten zijn om meerdere redenen interessant.

In de eerste plaats omdat onderzoeken zoals die van Jencks en Fägerlind (1975) belangrijk blijken te zijn. Zowel de pers als de mensen die het onderwijsbeleid maken, hebben er veel aandacht aan besteed.

In de tweede plaats is het interessant te weten in hoeverre het model en de resultaten van Jencks en Fägerlind gegeneraliseerd kunnen worden naar de Nederlandse situatie. Het heeft weinig zin meer om de conclusies van Jencks te gebruiken in de discussie rond de contourennota nu de conclusies van D&dJ bekend zijn.

En in de derde plaats zijn er ongetwijfeld mensen die de concrete schattingen van pad- en determinatie-coëfficiënten in het artikel van D&dJ interessant vinden.

Met deze drie redenen corresponderen de volgende drie voorbehouden, die ik in de loop van dit artikel aannemelijk zal proberen te maken.

In de eerste plaats is het politieke belang van dit soort studies aanzienlijk groter dan het wetenschappelijke belang. Of, om het wat onvriendelijker te zeggen, de wetenschappelijke waarde van dit soort studies is uitermate dubieus.

In de tweede plaats heeft het, om dezelfde redenen, ook weinig zin om de conclusies van D&dJ te gebruiken in de discussies over onderwijsbeleid.

En in de derde plaats is er onvoldoende bewijs om de door D&dJ gepresenteerde schattingen serieus te nemen, hoe interessant ze ook mogen zijn.

Het artikel van D&dJ heeft, zoals gezegd, voornamelijk een replicatie-karakter. Daardoor maken ze zichzelf in zekere zin methodologisch onkwetsbaar. Critici worden expliciet doorverwezen naar Jencks. 'Indien men daarom Jencks' resultaten (en ook de onze) zou verwerpen, kunnen wij dat voor een groot deel begrijpen. Vindt men Jencks' theoretische onderbouwing niet ver-

werpeijk, dan kan men ons model en de daarmee verkregen Nederlandse resultaten ook niet verwerpen' (bladzijde 22). Deze vorm van argumentatie is misschien wat ongebruikelijk, maar lijkt toch volkomen legitiem. Men stelt zich op het standpunt van de tegenstander en laat zien dat zijn conclusies niet noodzakelijk uit zijn uitgangspunten volgen (hoewel men het met die uitgangspunten ook al niet eens is). Dit is wat Overton (1973) 'within-paradigm criticism' noemt. De geschiedenis van de wetenschap laat zien dat dit soort kritiek dikwijls effectiever is dan 'between-paradigm criticism', wat maar al te dikwijls leidt tot oeverloze en steeds terugkerende debatten over grondslagen. D&dJ tonen aan dat Jencks' conclusies populatie afhankelijk zijn, wat een belangrijke relativering is. Op dezelfde manier proberen Jaspars en De Leeuw (1977) aan te tonen dat de conclusies van Jensen en Eysenck over de invloed van genotypische intelligentie sterk afhankelijk zijn van de keuze van de test, de populatie, en het specifieke pad model.

D&dJ zijn overigens wat ambivalent over die uitgangspunten 'Wij zijn daarom van mening dat een *voorzichtig* gebruik van Jencks' en onze onderzoeksresultaten toegestaan is' (bladzijde 21). Vooral in hoofdstuk 4, resultaten, laten D&dJ zich echter op de gebruikelijke wijze meeslepen door het model. Het wemelt daar van de gevonden causale relaties, die zonder enig voorbehoud gepresenteerd worden. Het moet D&dJ toch opgevallen zijn, toen ze de literatuur voor hun SISWO studiebijeenkomst over 'Inequality' doorkeken, dat juist dit soort hoogst aanvechtbare uitspraken in de pers en in de beleidsnota's terecht komen. De vage theoretische bezwaren in de inleiding, de methodologische beperkingen in voetnoot 7, en de onderstreping van het woord 'voorzichtig', zullen wellicht aan de aandacht van de op concrete resultaten beluste lezer ontsnappen.

Het lijkt daarom noodzakelijk om wat dieper in te gaan op de vele bezwaren die men tegen Jencks (en dus ook tegen D&dJ) kan inbrengen. En dit eens te meer omdat de causale modellen als paddestoelen uit de grond springen (Shea, 1976, bespreekt er een aantal), terwijl de fundamentele methodologische beperkingen van dit soort modellen wat genegeerd worden. De computer draait wel, SPSS is gewillig, maar de *voorzichtigheid* is zo nu en dan ook bij D&dJ ver te zoeken.

1 Causale modellen

1.1 Causaliteit

Toen pad analyse rond 1920 voor het eerst door Wright werd beschreven, begon er direct al een nogal heftige discussie rond het begrip 'causaliteit', zoals dat in verband met pad analyse gebruikt werd (Wright 1921, 1923, 1934; Niles, 1922, 1923; Tukey, 1954). We gaan hier wat nader op in. Het gaat hierbij

niet om de rol die het causaliteitsbeginsel in het algemeen in de wetenschap speelt (zie hiervoor bv. Franck 1932, Lindsay en Margenau, 1936, hfdst. X, Lenzen 1954, Bunge 1957), maar om de definitie van causale relaties en hun verband met empirische, statistische gegevens.

In de hoogtijdagen van logisch positivisme en logisch empirisme werd het belang van causaliteitsrelaties in de wetenschappen bestreden door ondermeer Russell (1912) en Pearson (1912). Russell beweerde dat causale relaties van de vorm 'als A dan altijd B' niet of nauwelijks in de wetenschap voorkomen, en dat wetenschappelijke wetten kwantitatieve functionele relaties zijn. Filosofen hebben ten onrechte de nadruk op causaliteit gelegd, waarschijnlijk omdat ze het begrip functie niet beheersten. Pearson ging nog verder. De fundamentele relatie in de wetenschap is correlatie of contingentie, causaliteit is slechts het hypothetische speciale geval van perfecte correlatie (lineair of niet-lineair). Men is het er tegenwoordig over eens dat de kritiek van Pearson, zoals de meeste positivistische en empiristische kritiek, te ver ging en allerlei schadelijke beperkingen aan wetenschapsbeoefening oplegde. 'Even if the time of modern physics and evolution had failed to be so one-sided and causal, it would still be easy to see that theories with causal elements have been of the greatest importance in all quantitative or semiquantitative branches of science' (Tukey, 1954, p. 40). 'Now that we can see causation in its present role as a guide and framework for functional analysis, we are sure that it is a useful concept, and we see little relation to correlation coefficients' (idem, p. 37). Theorieën of modellen hebben dus causale componenten, en functionele relaties in een model kunnen causaal geïnterpreteerd worden. Causaliteitsregels zijn specifieke hypothesen binnen een mathematisch of statistisch model. Wright is op dit punt altijd zeer duidelijk geweest. 'The method depends on the combination of knowledge of the degree of correlation among the variables in a system with such knowledge as may be possessed of the causal relations. In cases in which the causal relations are uncertain the method can be used to find the logical consequences of any particular hypotheses in regard to them'. (Wright, 1921, p. 557). Hetzelfde standpunt neemt ook Dempster (1971, sectie 5.2) in. Wold (1969, p. 449) definieert: 'Cause-effect relationships are stimulus-response relationships in genuine or fictitious experiments'. Ook bij Wold is de stimulus-response relatie een functionele, en ook bij Wold is de causale relatie alleen gedefinieerd binnen het model.

Als er dus geen pijl is in het padmodel tussen variabele X en variabele Y, of als de geschatte pad-coëfficiënt niet significant van nul afwijkt, dan betekent dit *niet* dat X geen invloed heeft op Y. Deze conclusie is model-afhankelijk, als we het model veranderen kan de conclusie veranderen. We hebben dus vertrouwen in de conclusie naarmate het model bewezen heeft te werken in vergelijkbare situaties, of naarmate alternatieve specificaties van het model tot

dezelfde conclusie leiden. Met behulp van de statistiek kunnen we proberen een model of aspecten van een model te falsifiëren. Natuurlijk bewijst het mislukken van zo'n falsificatiepoging niet dat het model waar is. 'Negative support is unfortunately the best available, but accumulation of such support from many data sets eventually creates a sense of confidence akin to positive support. This is surely as close to proof as is logically possible with empirical phenomena'. (Dempster, 1971, p. 338). Het zal duidelijk zijn dat het onderzoek van Jencks, en de replicatie van D&dJ, niet veel 'sense of confidence' inspireren. Er worden vrijwel geen alternatieve causale modellen onderzocht, er worden geen variabelen gemanipuleerd, het model in zijn geheel wordt niet of nauwelijks getoetst. De causaliteits-conclusies zijn dus noodzakelijkerwijs van uiterst beperkte waarde.

1.2 Descriptief of functioneel

In dit verband lijkt het ook nuttig in te gaan op een onderscheid gemaakt door Tukey (1954, p. 36-38). Methoden om gegevens te analyseren kunnen descriptief, tangentiaal, en functioneel zijn. Descriptieve methoden beschrijven hooguit de populatie waaruit de steekproef getrokken is, of misschien alleen maar de steekproef zelf. Tangentiale methoden laten zien wat er gebeurt wanneer we kleine veranderingen in de onafhankelijke variabelen toelaten, en functionele methoden leiden tot meer algemene uitspraken, die ook grotere veranderingen toelaten. Alleen functionele methoden leiden tot causale modellen in strikte zin (vergelijk de definitie van Wold). Padanalyse kan op alle drie de niveaus gebruikt worden, en men moet zich terdege realiseren welk niveau men denkt te bereiken. Het lijkt mij dat toepassingen zoals die van Jencks aan de descriptieve kant van de schaal liggen, waarbij het inderdaad nog niet eens duidelijk is of zij de populatie of de steekproef beschrijven. De uitspraken en conclusies van Jencks zijn echter geformuleerd op functioneel niveau. D&dJ laten zien dat de conclusies binnen het model populatie afhankelijk zijn, waarmee ze aantonen dat de methode van Jencks hoogstens tangentiaal is. Ook de lineariteit van het padmodel wijst in deze richting, als de veranderingen klein zijn kunnen we de functionele relaties lineair veronderstellen.

1.3 Verdere beperkingen

We hebben gezien dat de causale conclusies die op basis van een analyse als die van Jencks of D&dJ getrokken kunnen worden van zeer beperkte waarde zijn, omdat hun invariantie onder populatie of modelkeuze op geen enkele manier aangetoond is. Als men in het model gelooft, volgen de conclusies vanzelf. Als men niet in het model gelooft, hebben de conclusies geen enkele waarde. En Jencks voert uitermate weinig redenen aan, waarom we eigenlijk in zijn specifieke padmodel moeten geloven (dit geldt trouwens voor het me-

rendeel van de sociale wetenschappelijke toepassingen van pad modellen, dit in tegenstelling tot toepassingen in de genetica waar het padmodel uit Mendeliaanse assumpties volgt). In hun voetnoot 7 noemen D&DJ nog een aantal andere beperkingen van padmodellen. Deze beperkingen komen ook terug in een nogal agressief artikel van Taylor (1973), die er terecht op wijst dat al deze beperkingen tezamen genomen noodzakelijkerwijs leiden tot een onderschatting van het effect van onderwijs op inkomen en beroepsniveau. Methodologisch gezien is het geen enkel probleem interactie variabelen op te nemen, te werken met dummy indicatoren voor niet-lineaire codering, en niet-recursiviteit toe te laten. Om het model ook maar enigszins aannemelijk te maken zijn al dit soort uitbreidingen nodig, of is het althans nodig om na te gaan welke effecten die uitbreidingen hebben. Jencks laat dit na, en hij laat daarom weer een kans voorbij gaan om de invariantie van zijn resultaten (en dus de geldigheid van zijn conclusies) aan te tonen.

2 De variabelen

2.1 *Variabelen en indicatoren*

Een van de meest in het oog vallende eigenschappen van studies gebaseerd op pad analyse is dat de hypothetische constructen waarin men uiteindelijk geïnteresseerd is voortdurend verward worden met de variabelen die geacht worden de constructen te meten. Als men geïnteresseerd is in de overerfbaarheid van intelligentie gebruikt men als indicator van intelligentie het IQ. Alle uitspraken op grond van deze analyse zouden nu over IQ moeten gaan, maar in feite blijken ze vrijwel altijd over intelligentie te gaan. Hetzelfde geldt voor kwaliteit van het gevolgde onderwijs, beroepsstatus, en zelfs voor variabelen als gezinsgrootte en inkomen. Er is veel additionele research nodig voor we kunnen aannemen dat de gekozen indicatoren inderdaad de best mogelijke zijn, en voor we kunnen aannemelijk maken dat de indicatoren de theoretische constructen inderdaad meten. In het geval van het onderzoek van Jencks is er geen sprake van dit soort additionele research, en toch worden de conclusies geformuleerd in termen van de constructen.

2.2 *Transformaties*

In de econometrie wordt over het algemeen de logaritme van het inkomen bestudeerd, omdat de verdeling van inkomens nog steeds zeer scheef is. Conclusies over effect van onderwijs op inkomen zijn dus in feite conclusies over het effect van onderwijs op log inkomen. Ook in dit geval zal op basis van aanvullend onderzoek aangetoond moeten worden, dat dit soort transformaties al dan niet tot betere meting van het theoretische construct leiden. Daarom nemen we bijvoorbeeld niet de wortel uit het aantal schooljaren, of het

kwadraat van de gezinsgrootte. De in 1.3 al genoemde mogelijke niet-lineaire coderingen, of niet-metrische versies van pad analyse, kunnen hierbij nuttig zijn.

3 De gegevens

3.1 Kwaliteit

Er is veel te doen geweest over de vraag of men gegevens uit verschillende onderzoeken wel mag combineren om ze te gebruiken in een enkele analyse. Het lijkt mij dat methodologische verbodsbepalingen langzamerhand uit de tijd raken. Alles mag, alleen moet men natuurlijk wel laten zien dat de conclusies die men trekt inderdaad uit de gedane bewerkingen volgen. In dit geval willen we uitspraken doen over de parameters van een multivariate verdeling. Op basis van aselechte steekproeven kunnen we deze parameters schatten. Dit betekent dus dat D&dJ van de impliciete veronderstelling uitgaan dat de steekproef in het nationaal verkiezingsonderzoek 1971, de GIT normeringssteekproef 1964, de 10.000 recruten uit 1974, en de ITS steekproef van schoolverlaters uit 1965, allemaal aselechte steekproeven uit dezelfde populatie zijn. Deze aanname lijkt onhoudbaar. Dit relativeert de waarde van de conclusies nog wat meer. Natuurlijk is het zo dat er geen betere gegevens zijn. Het lijkt echter aanzienlijk nuttiger om ervoor te gaan zorgen dat ze er wel komen. Uit het oogpunt van eenvoudige wetenschappelijke kwaliteitscontrole is het betreurenswaardig dat analyses gepubliceerd worden waarvan niet eens de mate van misleidendheid vastgesteld kan worden.

3.2 Correlaties

De gegevens worden over het algemeen in de vorm van correlaties weergegeven. Dit is betreurenswaardig, omdat Tukey (1954) en Turner en Stevens (1959) hebben laten zien dat het over het algemeen zinvoller is varianties en covarianties te analyseren. Doordat correlaties sterk afhankelijk zijn van variantiefluctuaties zijn ze minder invariant, wat als gevolg heeft dat de analyse meer descriptief en minder functioneel wordt. Wright (1960) heeft laten zien dat het voldoende is om de varianties te vermelden, men kan dan de resultaten altijd nog hernormaliseren. In het onderzoek van D&dJ zou het bijvoorbeeld interessant zijn varianties bij mannen en vrouwen te vergelijken.

4 De bewerkingen

4.1 Regressie

D&dJ gebruiken multiële regressie om de pad coëfficiënten te schatten. Dat is in zekere zin jammer. Jencks gebruikt, in navolging van Wright, manipu-

laties van de algebraïsche vergelijkingen die uit de gereduceerde vorm van het model volgen. Dit lijkt beter, omdat het expliciet rekening houdt met de diverse restricties die het model impliceert. Er zijn trouwens, voornamelijk in de econometrie, aanzienlijk betere methoden ontwikkeld (een goed recent overzicht is Dijkhuizen, 1977). Programma's om efficiënte schatters te berekenen zijn eveneens verkrijgbaar (Jöreskog, 1973).

Aan de andere kant zal de schattingsmethode in dit geval niet veel invloed hebben. Het model van D&dJ is namelijk niet erg restrictief. In tabel 3 en tabel 4 vinden we 44 padcoëfficiënten. Tezamen met de correlaties tussen de drie exogene variabelen levert dit 47 parameters op. Als we de correlaties in tabel 1 en 2 tellen vinden we 48 te fitten correlaties. Ook uit deze eenvoudige telling blijkt de descriptieve aard van het model, vrijwel ieder stelsel van 48 correlaties zal in het model te passen zijn. In principe zou men natuurlijk die ene overblijvende vrijheidsgraad kunnen gebruiken om het model te toetsen, maar dit lijkt een nogal hopeloze onderneming.

4.2 *Standaardfouten*

Ernstiger is dat bij een nauwelijks geïdentificeerd model zoals dit de schattingen van de padcoëfficiënten nooit erg precies kunnen zijn. We hebben overidentificatie nodig om de standaardfouten kleiner te maken. De gebruikte techniek om standaardfouten te berekenen is nogal misleidend, omdat de normaliteit van de variabelen aangenomen wordt. Met name voor gezinsgrootte, inkomen, en genoten onderwijs doet deze aanname wat vreemd aan. Bovendien is het toetsen van individuele regressiecoëfficiënten misleidend wanneer de schattingen van de regressiecoëfficiënten gecorrigeerd zijn. Beter is weer een hiërarchische analyse, waar paden om de beurt weggelaten worden. Dit geeft ook inzicht in de mate waarin multicollineariteit een rol speelt.

5 Samenvatting

Ons belangrijkste uitgangspunt is dat het mogelijk is causale relaties aannemelijk te maken op basis van statistische analyse van empirische gegevens. Pad modellen zijn hiervoor in principe geschikt. Maar beslist nodig is dat veel verschillende causale hypothesen uitgetest worden, dat het effect van verschillende restrictieve assumpties nagegaan wordt, dat de resultaten bij onafhankelijke replicaties vergelijkbaar blijken te zijn, en dat waar mogelijk de determinerende variabelen gemanipuleerd worden. En dit geldt met name in de exploratieve fase van onderzoek, waarin er nog geen definitief, goed gecorroborerd model bestaat. Als aan deze voorwaarden niet wordt voldaan hebben pad modellen slechts een zuiver descriptieve waarde. Gebruik van causaliteitsterminologie, en verwisseling van constructen met hun imperfecte

operationalisaties, werkt dan uiterst misleidend.

Het onderzoek van Jencks (en daardoor ook dat van D&dJ) is volgens deze criteria zuiver discriptief. Niettemin worden (vooral bij Jencks) de resultaten in functionele causaliteitsterminologie beschreven, en worden er zelfs politieke aanbevelingen op basis van deze resultaten gedaan. Dit is misbruik maken van wetenschappelijke methoden, omdat het een precisie en invariantie suggereert die op geen enkele manier aangetoond is. De povere kwaliteit van de gebruikte gegevens en de statistische beperkingen en tekortkomingen van dit soort onderzoek versterken deze conclusie alleen nog maar.

Niettemin is het onderzoek van D&dJ nuttig, ja zelfs zeer nuttig. Ze hebben zonneklaar aangetoond dat de resultaten van Jencks in Nederland onbruikbaar zijn. Voorlopig geldt datzelfde helaas ook nog voor hun eigen resultaten, maar de verzamelde gegevens kunnen verder verwerkt, aangevuld, en verbeterd worden. Misschien kunnen dan te zijner tijd de resultaten gebruikt worden bij het maken (en niet bij het rationaliseren) van beleidsbeslissingen. Ik denk dat het moeilijk zal zijn om tot een model te komen dat in diverse interessante situaties invariante resultaten geeft. Maar het is het proberen waard.

Literatuur

- Bunge, M., *Causality*. Harvard University Press, Cambridge, Mass., 1957.
- Dempster, A. P., An overview of multivariate data analysis. *J. Multivariate analysis*, 1, 1971, 316-346.
- Dijkhuizen, A. A., *Klassieke econometrische methoden voor het schatten van de structuurparameters in het padanalytische model*. Rotterdam, Projekt Onderwijs en Sociaal Milieu. Mededeling no. 23, 1977.
- Fägerlind, I., *Formal education and adult earnings*. Almqvist & Wicksell, Stockholm, 1975.
- Frank, P., *Das Kausalgesetz und seine Grenzen*. Springer, Wenen, 1932.
- Jaspars, J. M. F., & De Leeuw, J., *Models of psychometrical genetics*. Voordracht Onderwijsresearch congres, Leiden, 1977.
- Jencks, C., e.a., *Inequality*. Basic Books, New York, 1972.
- Jöreskog, K., A general method for estimating a linear structural equation system. In: *Structural equation models in the social sciences*. Seminar Press, New York 1973.
- Lenzen, V. F., *Causality in natural science*. Thomas, Springfield, 1954.
- Lindsay, R. B., & Margenau, H., *Foundations of physics*. Dover, New York, 1957, (heruitgave eerste druk 1936).
- Niles, H. E., Correlation, causation, and Wright's theory of 'path coefficients'. *Genetics*, 7, 1922, 258-273.
- Niles, H. E., The method of path coefficients: an answer to Wright. *Genetics*, 8, 1923, 256-260.
- Overton, W. F., On the assumptive base of the nature-nurture controversy: additive versus interactive conceptions. *Human Develop*, 16, 1973, 74-89.

- Pearson, K., *The Grammar of Science*. Smith, Gloucester, 1957 (heruitgave derde druk 1911).
- Russell, B., On the notion of cause. In: *Mysticism and logic*, Unwin, London, 1963. (from Proc. Aristotelian Soc, 1912).
- Shea, B. M., Schooling and its antecedents: substantive and methodological issues in the status attainment process. *Rev. of Educ. Research*, 46, 1976, 463-526.
- Taylor, H. F., Playing the dozens with path analysis: methodological pitfalls in in Jencks et al: 'Inequality'. *Sociology of Educ.*, 46, 1973, 433-450.
- Tukey, J. W., Causation, regression, and path analysis. In: *Statistics and mathematics in biology*. Iowa State University Press, Ames, 1954.
- Turner, M. E., & Stevens, C. E., The regression analysis of causal paths. *Biometrics*, 15, 1959, 236-258.
- Wold, H. O., Mergers of economics and philosophy of science. *Synthese*, 20, 1969, 427-482.
- Wright, S., Correlation and causation. *J. Agricultural research*, 20, 1921, 557-585.
- Wright, S., The theory of path coefficients: a reply to Niles's criticism. *Genetics*, 8, 1923, 239-255.
- Wright, S., The method of path coefficients. *Ann. Math. Statist.*, 5, 1934, 161-215.
- Wright, S., Path coefficients and path regressions: alternative of complementary concepts? *Biometrics*, 16, 1960, 189-202.

Weerwoord op De Leeuw 'Sociale wetenschappen tussen relevantie en methode'

J. Dronkers en U. de Jong

De methodologische bezwaren van De Leeuw tegen het onderzoek van Jencks (en dus tegen onze replicatie) die hij in 1 (causale modellen), 2 (De variabelen), 3 (De gegevens) en 4 (De bewerkingen) opsomt, delen wij. In onze kritiek op Jencks hebben wij niet uitsluitend de weg van De Leeuw gekozen, nl. de opsomming van methodologische bezwaren die elders reeds geuit zijn. Naast ons bibliografisch overzicht van de discussie naar aanleiding van Jencks (De Jong en Dronkers, 1977) kozen wij ook een andere, wellicht minder gebruikelijke weg, nl. het demonstreren van Jencks werkwijze voor het Nederlands publiek zodat ieder de ernstige beperkingen van die werkwijze duidelijk kan zien. Heel bewust hebben wij niet meer gedaan dan Jencks in het door ons geciteerde deel van zijn Appendix B deed. Wij hebben dus geen alternatieve causale modellen onderzocht, geen interactie tussen de variabelen trachten te bepalen, etc. De resultaten, die wij in hoofdstuk 4 van ons artikel presenteren, passen binnen deze doelstelling. Wij laten daar zien welke conclusies er op