

Causale analyse van paneldata*

J. Dessens en W. Jansen**

Inleiding

Aan sociologiestudenten wordt altijd voorgehouden, dat het experiment de meest geëigende grondvorm is ter toetsing van causale hypothesen; immers, in het experiment kunnen we het effect van de onafhankelijke variabele vaststellen, dankzij het feit dat we de onafhankelijke variabele zelf kunnen variëren en de mogelijke storende effecten onder controle kunnen houden middels randomisings- en matchings-procedures. Er wordt echter onmiddellijk aan toegevoegd, dat experimenten binnen de sociale wetenschappen moeilijk te realiseren zijn, aangezien veel variabelen binnen het sociologisch onderzoek niet naar believen kunnen worden gevarieerd, laat staan dat onderzoeksgroepen gerandomiseerd dan wel gematched zouden kunnen worden. Niettemin kan het experiment worden beschouwd als dé grondvorm van onderzoek. Het experiment heeft voor alle andere grondvormen model gestaan, zodat bij elke grondvorm in meer of mindere mate elementen van het experiment zijn terug te vinden. Zo is in het kader van de één-moment-survey veel energie gestoken in het benaderen van een aantal specifieke elementen uit het experiment (hierbij valt bijvoorbeeld te denken aan het matchen uit de experimentele situatie en de bekende elaboratie-procedure van Lazarsfeld, die in feite niets anders is dan het achteraf 'matchen' van de onderzoeksgroep op mogelijke interfererende variabelen).

Maken we de balans op van wat er in feite in een survey ge-

* Wij danken Dr. P. G. Swanborn voor zijn waardevolle commentaar op een eerdere versie van dit artikel.

** De auteurs studeerden af aan de Rijksuniversiteit Utrecht en zijn thans als wetenschappelijk medewerker werkzaam binnen de vakgroep Theorie en Methodologie van het Sociologisch Instituut Utrecht.

beurt, dan komen we tot het volgende, sterk vereenvoudigde schema:

1. meting op één tijdstip van alle voor de probleemstelling relevant geachte variabelen (inclusief mogelijke interfererende variabelen);
2. het a priori ordenen van de variabelen op de tijdsdimensie;
3. het achteraf 'matchen' op interfererende variabelen.

Problematisch is het onder 2. gestelde: vaststelling van de tijdsvolgorde en daarmee de richting van causale beïnvloeding tussen de variabelen. Men is, indien geconfronteerd met dit probleem geneigd het roer om te gooien en over te gaan op méér-moment-opnamen i. c. panelstudies. Wij zijn echter met Gadourek (1974) van mening, dat de keuze voor een grondvorm als het méér-moment-survey afhankelijk zal moeten zijn van de probleemstelling. Wil men bijvoorbeeld de invloed van bepaalde waarden en normen op een attitude vaststellen, dan zal men, indien deze waarden en normen sterk tijdgebonden zijn, een andere grondvorm dan het méér-moment-survey dienen te kiezen, b. v. meting op één tijdstip bij verschillende generaties. Algemeen, indien de inhoud van de variabelen varieert naar hun relatieve positie op de tijds- en/of ruimte-dimensie dienen we daarmee bij de keuze van de grondvorm van onderzoek rekening te houden. Een voorbeeld is de variabele 'religie'. De inhoud van deze variabele is in de tijd en ook naar de plaats aan behoorlijke verandering onderhevig. Vinden we nu bij analyse van één-moment-survey gegevens een relatie tussen de variabele religie en een of andere gedragsvariabele, een relatie die bovendien varieert met een indeling naar leeftijdsgroepen of urbanisatiegraad, dan kunnen we op basis hiervan stellen dat binnen de categorieën van de variabele leeftijd of urbanisatiegraad het verband tussen religie en de gedragsvariabele gelijk blijft dan wel verschillend komt te liggen dan wel verdwijnt. We kunnen echter niets concluderen met betrekking tot de ontwikkeling in de tijd (of ruimte) van de relatie. Het maakt derhalve voor de keuze van de grondvorm verschil of we geïnteresseerd zijn in de relatie tussen religie en de gedragsvariabele zonder meer (hier zullen we een één-moment-survey kiezen) of dat we juist geïnteresseerd zijn in de ontwikkeling van de relatie in tijd en/of ruimte (hier zullen we voor een méér-moment-survey kiezen). Dat bovenstaande geenszins nieuw is (al kan er niet genoeg de nadruk op worden gelegd) blijkt ook wel uit de meer formele systematisering van Galtung (1970) van het bovenstaande.

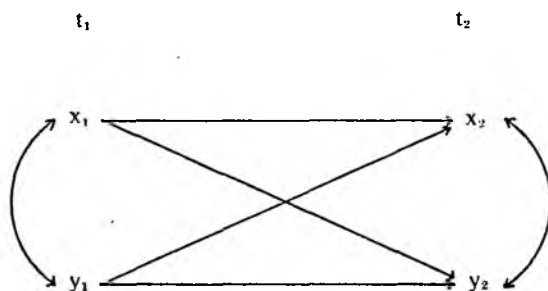
In het navolgende zullen we ons meer expliciet bezig houden met

het méér-moment-survey, hierna zoals algemeen gebruikelijk aan te duiden met de term panelanalyse. We volgen het historische proces volgens hetwelk men met behulp van panelanalyse causale relaties stevig in de vingers probeerde te krijgen. Achtereenvolgens komen hierbij aan de orde de zogenaamde 'cross-lagged panel correlation technique' van Campbell en de zogenaamde 'sixteen-fold table technique' van Lazarsfeld. Daarna worden deze technieken met elkaar vergeleken en enkele aanvullingen en correcties van Yee en Gage en Rozelle en Campbell genoemd. Tenslotte behandelen we de padanalyse als analyse-strategie bij paneldata, welke analyse-procedure tevens extra toetsingsmogelijkheden geeft b.v. in het geval van gecorreleerde restvariabelen. In de tweede helft van dit artikel gaan we daar uitgebreider op in.

Twee analysestrategieën

Van de verschillende analysestrategieën om de tijdsvolgorde van twee variabelen vast te stellen bij panelanalyse zijn de twee meest bekende die van Campbell, ontwikkeld vanuit de sociaal-psychologische experimentele hoek (Campbell, 1963), en die van Lazarsfeld, ontwikkeld vanuit de in de sociologie bestaande surveytraditie (Lazarsfeld, 1972). Beide technieken zijn gebaseerd op het volgende grondschema:

Figuur 1



waarin: X_1 en Y_1 de op tijdstip t_1 gemeten variabelen voorstellen en X_2 en Y_2 dezelfde variabelen, op tijdstip t_2 gemeten.

De door Campbell ontwikkelde 'cross-lagged panel correlation technique' komt neer op de, vanuit de experimentele onderzoeks-

opzet gezien, plausibele onderlinge vergelijking van de correlaties $r_{X_1 Y_2}$ en $r_{Y_1 X_2}$, waarbij geldt:

- indien $r_{X_1 Y_2} > r_{Y_1 X_2}$ dan is het effect van X op Y groter dan het effect van Y op X;
- indien $r_{X_1 Y_2} < r_{Y_1 X_2}$ dan is het effect van Y op X groter dan het effect van X op Y.

Als illustratief voorbeeld hanteren we in dit artikel de bekende paneldata van Lipset e. a. (Lipset, 1954), waarbij de variabele X overeenkomt met stemintentie (republikeins vs. democratisch) en de variabele Y overeenkomt met de houding t.o.v. de republikeinse presidentskandidaat Wilkie (positief vs. negatief), op twee achtereenvolgende tijdstippen gemeten tijdens de presidentsverkiezingscampagne in de V.S. in 1940.

De empirische correlaties $r_{X_1 Y_2}$ en $r_{Y_1 X_2}$ zijn respectievelijk .65 en .51.

Met behulp van de 'cross-lagged panel correlatie techniek' besluiten we op grond van deze correlaties, dat stemintentie eerder tot een al dan niet positieve houding t.o.v. de presidentskandidaat zal leiden dan andersom.

De door Lazarsfeld ontwikkelde techniek om de causale richting tussen twee variabelen te toetsen staat bekend als de '16-fold-table-technique'. Deze 16-cels tabel ziet er als volgt uit:

Tabel 1

		2e meting			
1e meting		$X_2^- Y_2^-$	$X_2^- Y_2^+$	$X_2^+ Y_2^-$	$X_2^+ Y_2^+$
X_1^-	Y_1^-	(1, 1)	(1, 2)	(1, 3)	(1, 4)
X_1^-	Y_1^+	(2, 1)	(2, 2)	(2, 3)	(2, 4)
X_1^+	Y_1^-	(3, 1)	(3, 2)	(3, 3)	(3, 4)
X_1^+	Y_1^+	(4, 1)	(4, 2)	(4, 3)	(4, 4)

De eenheden, die tussen beide metingen op één variabele zijn veranderd in congruente richting (d.w.z. stemintentie en houding t.o.v. presidentskandidaat zijn met elkaar in overeenstemming

gebracht) zijn te vinden in de cellen (2, 1), (2, 4), (3, 1) en (3, 4).

Indien de variabele X een positief effect op variabele Y zou hebben, dan dient te gelden: $(2, 1) + (3, 4) > (3, 1) + (2, 4)$. In woorden: het patroon - + wordt vaker veranderd in -- dan in ++ en het patroon + - vaker in ++ dan in +-. De 'gevolg'-variabele Y verandert sneller dan de relatief stabielere 'oorzaak'-variabele X.

Stellen we op grond van de cellen (2, 1), (2, 4), (3, 1) en (3, 4), die overeenkomen met veranderingen in congruente richting, de volgende tabel op:

Tabel 2

(2, 1)	(2, 4)	(2, 1) + (2, 4)
(3, 1)	(3, 4)	(3, 1) + (3, 4)
(2, 1) + (3, 1)	(2, 4) + (3, 4)	N_K

In het geval, dat de variabelen X en Y statistisch onafhankelijk zijn is het aantal verwachte veranderingen in voor X congruente richting met betrekking tot cel (2, 1) gelijk aan:

$$((2, 1) + (2, 4))((2, 1) + (3, 1))/N_K = A$$

Voor de cel (3, 4) krijgen we:

$$((3, 1) + (3, 4))((2, 4) + (3, 4))/N_K = B$$

Het totale aantal boven kans voorkomende veranderingen in voor X congruente richting is derhalve:

$$((2, 1) - A) + ((3, 4) - B) = 2((2, 1)(3, 4) - (2, 4)(3, 1))/N_K$$

Definiëren we $((2, 1)(3, 4) - (2, 4)(3, 1))$ als dK , dan gaat de uitdrukking over in $2dK/N_K$.

De eenheden, die tussen beide metingen op één variabele in incongruente richting zijn veranderd (d.w.z. stemintentie en houding t.o.v. presidentskandidaat zijn niet meer met elkaar in overeenstemming) zijn te vinden in de cellen (1, 2), (1, 3), (4, 2) en (4, 3). Nog steeds onder de hypothese dat de variabele X een positief effect op de variabele Y heeft, dient hier te gelden: $((1, 2) + (4, 3)) < ((1, 3) + (4, 2))$. In woorden: het patroon -- wordt vaker veranderd in + - dan in - + en het patroon + + vaker in - + dan in + -.

Analoog aan het congruente geval kunnen we op basis van de cellen, die overeenkomen met verandering in incongruente richting de volgende tabel opstellen:

Errata artikel: 'Causale analyse van paneldata', Mens en maatschappij, 51, no. 1, 1976, blz. 2-25.

Blz. 7; regel 16/17

'tegengesteld' vervangen door 'gelijk'.

Blz. 7; regel 21/22

Bij beide N_K 's en N_I 's zijn de kwadraten weggevallen.

Blz. 7; regel 27 e.v.

De zin: 'Om deze reden . . .' vervangen door: -

Hieruit volgt dat
$$\frac{-N_K}{4} \leq \frac{dK}{N_K} \leq \frac{N_K}{4} \quad \text{en} \quad \frac{-N_I}{4} \leq \frac{dI}{N_I} \leq \frac{N_I}{4},$$

zodat
$$\frac{-(N_K + N_I)}{4} \leq \frac{dK}{N_K} + \frac{dI}{N_I} \leq \frac{(N_K + N_I)}{4}. \quad \text{Aangezien}$$

$N_K + N_I$ maximaal gelijk is aan N , zal gelden:
$$-N/4 \leq \frac{dK}{N_K} + \frac{dI}{N_I} \leq N/4.$$

Hieruit volgt: $-N/2 \leq O_{XY} \leq N/2.$

Blz. 7; laatste regel

$(N_K + N_I)/4$ vervangen door $N/2$.

Blz. 8; regel 3

In de formule voor O'_{XY} 8 vervangen door 4 en $N_K + N_I$ door N .

Toevoegen: Om onverklaarbare redenen normeert Lazarsfeld deze maat tussen -2 en $+2$, zodat in zijn formule in de teller het getal 8 voorkomt. Wij normeren O'_{XY} tussen -1 en $+1$.

Blz. 8; regel 15

De zinnen: ' Immers indien $q \dots$ ongevoelig hiervoor' vervangen door: Immers indien $p \Rightarrow q$, dan zal dK relatief klein (soms zelfs negatief) en dI relatief groot worden. Dit hangt samen met het feit, dat de verandering in X zoveel groter is dan de verandering in Y . Aangezien dK en dI in de formule van O'_{XY} worden gesommeerd, vallen deze effecten min of meer

tegen elkaar weg, zodat gesteld kan worden, dat O'_{XY} relatief ongevoelig is voor het verschil in turnover tussen X en Y.

Blz. 8; regel 25/26

Achter $(2 + 3) > (1 + 1)$ toevoegen: zodat dI negatief en dK positief is. In de formule voor O'_{XY} 8 vervangen door 4 en .13 door .07.

Blz. 9; regel 17 na tabel

Tot en met blz. 10, regel 10, vervangen door:

De reden hiervan valt eenvoudig in te zien, indien we alle mogelijke effecten beschouwen:

1. X_1 heeft een positief effect op Y_2 ;
2. X_1 heeft een negatief effect op Y_2 ;
3. Y_1 heeft een positief effect op X_2 ;
4. Y_1 heeft een negatief effect op X_2 .

In de situatie dat alleen 1. en 3. of 2. en 4. voorkomen, kunnen we $r_{X_1Y_2}$ en $r_{Y_1X_2}$ zonder meer vergelijken om de dominante oorzaak vast te stellen. In de situatie waarin dit niet het geval is, b.v. bij aanwezigheid van naar één of enkele relevante kenmerken onderscheiden subgroepen binnen de totale onderzoeksgroep, kunnen alle vier effecten tegelijkertijd een rol spelen en is vergelijking van $r_{X_1Y_2}$ en $r_{Y_1X_2}$ weinig zinvol.

Blz. 18; regel 19

Het derde toetsingscriterium dient te vervallen.* De intervallen voor de waarden van p_{11} en p_{12} op blz. 19 veranderen hierdoor niet; de nieuwe intervallen voor de waarden van p_{31} en p_{32} op blz. 23 worden: $-.21 \leq p_{31} \leq .88$ en $.04 \leq p_{32} \leq .62$.

Blz. 23; regel 7

De conclusie blijft gehandhaafd, behalve bij een waarde voor $r_{1e_3} = .80$.

Blz. 23; bij opsomming

Toevoegen: 4. de schattingen van de parameters liggen binnen relatief kleine intervallen en de hypothese omtrent de dominante oorzaak wordt weerlegd.

Blz. 23; laatste regel

Toevoegen achter: '... het geval is': terwijl de vierde situatie het meest informatief is.

J. Dessens
W. Jansen

* Wij danken deze opmerking aan W. Saris.

Tabel 3

(1, 2)	(1, 3)	(1, 2) + (1, 3)
(4, 2)	(4, 3)	(4, 2) + (4, 3)
(1, 2) + (4, 2)	(1, 3) + (4, 3)	N_I

Het aantal beneden kans voorkomende veranderingen in voor X incongruente richting voor de cellen (1, 2) en (4, 3) is respectievelijk:

$$-(1, 2) + ((1, 2) + (1, 3))((1, 2) + (4, 2))/N_I \text{ en} \\ -(4, 3) + ((4, 2) + (4, 3))((1, 3) + (4, 3))/N_I$$

Uitgewerkt geeft dit voor de totale beneden kans voorkomende veranderingen in voor X incongruente richting:

$$2((1, 3)(4, 2) - (1, 2)(4, 3))/N_I$$

Definiëren we $((1, 3)(4, 2) - (1, 2)(4, 3)) = dI$, dan gaat de uitdrukking over in $2dI/N_I$.

Als maat voor de relatieve sterkte van de invloeden van X op Y en Y op X kunnen we nu de volgende formule hanteren:

$O_{XY} = 2(dK/N_K + dI/N_I)$, waarbij in het algemeen dK en dI te-gengesteld van teken zullen zijn. Indien nu O_{XY} positief is, dan is het effect van X op Y groter dan het effect van Y op X en indien O_{XY} negatief is, dan is het effect van Y op X groter dan het effect van X op Y.

Algebraïsch kan worden afgeleid, dat $-N_K/4 \leq dK \leq N_K/4$ en $-N_I/4 \leq dI \leq N_I/4$. Dit valt echter ook intuïtief eenvoudig in te zien: dK zal maximaal zijn indien de som van de cellen op de hoofddiagonaal gelijk is aan N_K en N_K zodanig over de cellen (2, 1) en (3, 4) is verdeeld, dat $(2, 1) = (3, 4) = N_K/2$.

dK is minimaal indien de som van de niet-diagonale cellen gelijk is aan N_K en $(2, 4) = (3, 1) = N_K/2$. Analooq voor dI . Om deze reden is de bovengrens van O_{XY} gelijk aan $(N_K + N_I)/4$ en de benedengrens $-(N_K + N_I)/4$.

Om O_{XY} nu onafhankelijk te maken van de absolute grootte van N_K en N_I delen we O_{XY} door $(N_K + N_I)/4$, met als resultaat de

door Lazarsfeld voorgestelde maat O'_{XY} voor de analyse van de causale richting tussen de variabelen X en Y:

$$O'_{XY} = \frac{8(\frac{dK}{N_K} + \frac{dI}{N_I})}{N_K + N_I}$$

Alvorens een toepassing te geven van deze techniek op de data van Lipset e. a., gaan we nader in op de grootheden dK en dI.

De totale turnover¹ van variabele X is gelijk aan:

$$((1, 3) + (1, 4) + (2, 3) + (2, 4) + (3, 1) + (3, 2) + (4, 1) + (4, 2))$$

en de totale turnover van variabele Y is gelijk aan:

$$((1, 2) + (1, 4) + (2, 1) + (2, 3) + (3, 2) + (3, 4) + (4, 1) + (4, 3)).$$

De 'overlap' is gelijk aan $((1, 4) + (2, 3) + (3, 2) + (4, 1))$. De turnover van X en Y, ontdaan van deze overlap is respectievelijk:

$$((1, 3) + (2, 4) + (3, 1) + (4, 2)) = p$$

$$((1, 2) + (2, 1) + (3, 4) + (4, 3)) = q$$

Indien p nu aanzienlijk groter is dan q (of omgekeerd) dan zal dit effect hebben op het teken van dK en dI. Immers indien q relatief groot wordt t. o. v. p en $dK > 0$, dan zal de kans dat dI eveneens groter is dan nul groter worden. Dit valt eenvoudig in te zien aan de hand van de formules voor dK en dI. In dit geval zijn dK en dI niet meer verschillend van teken. Aangezien het effect van dK en dI in tegengestelde richting werkt is O'_{XY} relatief ongevoelig hiervoor.

We passen nu de aldus verkregen maat toe op de data van Lipset e. a. (zie tabel 4).

Nu geldt: $(2, 1) + (3, 4) > (3, 1) + (2, 4)$, omdat $(11 + 11) > (1 + 1)$

en: $(1, 2) + (4, 3) > (1, 3) + (4, 2)$, omdat $(2 + 3) > (1 + 1)$.

$$O'_{XY} = \frac{8(\frac{120}{24} - \frac{5}{7})}{266} = .13$$

Uit bovenstaande resultaten volgt, dat het effect van de variabele stemintentie groter is dan het effect van de variabele houding t. o. v. Wilkie. De eerstgenoemde variabele dient daarom beschouwd te worden als de dominante oorzaak.

Tabel 4

2e meting

1e meting		St.	H.	St.	H.	St.	H.	St.	H.
stem- int.	houding t.o.v. Wilkie	Dem.	Neg.	Dem.	Pos.	Rep.	Neg.	Rep.	Pos.
Dem.	Neg.	68		2		1		1	
Dem.	Pos.	11		12		0		1	
Rep.	Neg.	1		0		23		11	
Rep.	Pos.	2		1		3		129	

Nu zijn zowel bij de 'cross-lagged panel correlation technique' als bij Lazarsfelds '16-fold-table technique' enige kanttekeningen te maken. Het voornaamste bezwaar ten aanzien van de '16-fold-table' is de beperking tot dichotome data. Yee en Gage (1968) geven m.b.v. hun zgn. FSM-techniek een uitbreiding tot getrichotomiseerde data - maar dit legt enige beperkingen op ten aanzien van het meetniveau - en tot continue data met behulp van hun zgn. FCP-techniek. Ten aanzien van elk der drie technieken blijft dan nog gelden, dat de analyse bij uitbreiding van het aantal variabelen nogal gecompliceerd wordt.

Ten aanzien van de 'cross-lagged panel correlation technique' maken zowel Rozelle en Campbell (1969) als Yee en Gage (1968) de opmerking dat een onderlinge vergelijking van $r_{X_1 Y_2}$ en $r_{Y_1 X_2}$

in lang niet alle gevallen uitsluitel kan geven omtrent de causale richting, m.a.w. welke variabele als oorzaak en welke als effect dient te worden aangemerkt.

De reden hiervan valt eenvoudig in te zien in het geval van variabelen, die elkaar wederzijds beïnvloeden. Indien we b.v. vinden, dat $r_{X_1 Y_2} > r_{Y_1 X_2}$ dan kunnen we niet zonder meer stel-

len, dat X de oorzaak is van Y, omdat we te maken hebben met meerdere effecten, die niet van elkaar te onderscheiden zijn.

Wanneer we b.v. de variabelen 'werkgroepbezoek' en 'studie-resultaat' op twee tijdstippen zouden meten, dan zou kunnen gelden, dat een hoge mate van werkgroepbezoek op tijdstip t_1 leidt tot goede studieresultaten op tijdstip t_2 en dat goede studieresul-

taten op tijdstip t_1 leiden tot een hoge mate van werkgroepabsenteïsme op tijdstip t_2 . Voor de correlatiecoëfficiënt $r_{Y_1 X_2}$ kunnen we eenzelfde redenering opzetten.

De oorzaak van deze samengestelde effecten kan gevonden worden in de aanwezigheid van naar één of enkele kenmerk(en) onderscheiden subgroepen binnen de totale onderzoeksgroep. Zo zal voor het gegeven voorbeeld gelden, dat het laatstgenoemde effect (goede studieresultaten leiden tot een hoge mate van absenteïsme) zal gelden voor de 'individualisten' in de totale onderzoeksgroep; het eerstgenoemde effect voor de 'collectivisten'. Opnemen in het onderzoek van een of meerdere kenmerk(en), waarop subgroepen verschillen is hier een vereiste; daarna kunnen binnen de afzonderlijke subgroepen de 'cross-lagged' correlatiecoëfficiënten zinvol vergeleken worden.

Dat dit probleem eerst in 1969 door Rozelle en Campbell werd gesignaleerd, hangt wellicht samen met het feit, dat twee van de vier mogelijk te assumeren effecten van te voren vaak als niet plausibel kunnen worden beschouwd en daarom niet in ogenschouw behoeven te worden genomen.

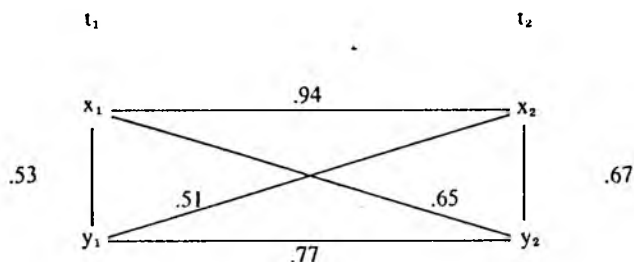
Directe en indirecte effecten

Noch de 'cross-lagged correlation technique' noch Lazarsfelds 'sixteen-fold table technique' houdt rekening met zogenaamde indirecte effecten, d.w.z. dat Y_1 via X_1 nog invloed heeft op X_2 .

Om nu een indruk te krijgen van het directe effect van Y_1 op X_2 zouden we de partiële correlatiecoëfficiënt $r_{X_2 Y_1 \cdot X_1}$ kunnen voorstellen. Het naar nul gaan van deze maat impliceert dat er geen direct effect van Y_1 op X_2 aanwezig is, maar een 'spurious' correlatie als gevolg van niet in het model opgenomen indirecte effecten, die leiden tot een correlatie tussen X_1 en Y_1 . Het bovenstaande geldt op analoge wijze voor de relatie tussen X_1 en Y_2 . Hier dienen we te corrigeren voor indirecte effecten via Y_1 .

Daarom ligt het voor de hand voortaan $r_{X_2 Y_1 \cdot X_1}$ en $r_{Y_2 X_1 \cdot Y_1}$ met elkaar te vergelijken teneinde de dominante oorzaak vast te stellen. Indien we het bovenstaande toepassen op ons eerder gegeven voorbeeld krijgen we:

Figuur 2



Zoals eerder vermeld is $r_{X_1 Y_2} > r_{Y_1 X_2}$; immers $.65 > .51$.

Vergelijken we nu de partiële correlatiecoëfficiënten $r_{X_1 Y_2 \cdot Y_1} = .52$ en $r_{X_2 Y_1 \cdot X_1} = .04$, dan zien we dat weliswaar de voorspel-

ling t.a.v. welke variabele als meest dominante oorzaak optreedt, dezelfde is, maar tevens zien we dat de relatie tussen X_2 en Y_1 vrijwel geheel 'spurious' is; informatie die we via de 'cross-lagged correlation technique' en Lazarsfelds 'sixteen-fold table technique' niet verkregen.

Zouden we in plaats van partiële correlatiecoëfficiënten de ongestandaardiseerde partiële regressiecoëfficiënten $b_{Y_2 X_1 \cdot Y_1}$ en $b_{X_2 Y_1 \cdot X_1}$ berekenen, dan valt op grond van Blalock (1964, blz.

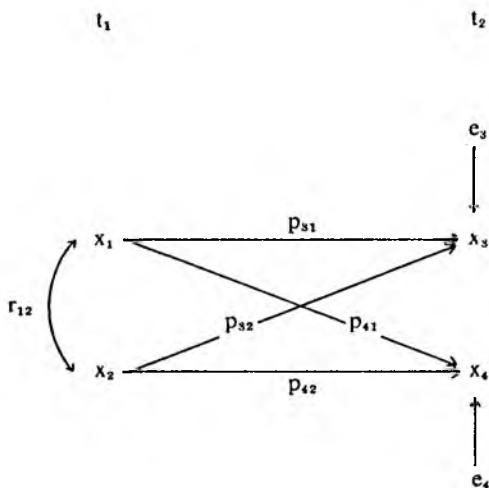
103, tabel 4) op te merken, dat bij vergelijking van verschillende steekproeven, waarin de varianties van de gemeten variabelen uiteen kunnen lopen, ongestandaardiseerde partiële regressiecoëfficiënten de voorkeur verdienen boven partiële correlatiecoëfficiënten: de eerste zijn relatief veel stabielier dan de laatste. Zijn we echter uitsluitend geïnteresseerd in de resultaten voor een bepaalde onderzoeksgroep, dan kunnen we zonder meer volstaan met partiële correlatiecoëfficiënten. Willen we echter de 'wetmatigheden' achterhalen, die ons in staat stellen voorspellingen te maken die ook voor andere onderzoeksgroepen kunnen gelden, dan dienen we gebruik te maken van ongestandaardiseerde partiële regressiecoëfficiënten.

Indien we iets willen zeggen over de relatieve grootte van de directe en indirecte effecten dan ligt het voor de hand padcoëfficiënten te gebruiken. Ter voorkoming van onnodig werken met

subscripts benoemen we de variabelen opnieuw en wel als volgt:
 $X_1 \rightarrow X_1$, $X_2 \rightarrow X_3$, $Y_1 \rightarrow X_2$ en $Y_2 \rightarrow X_4$.

Het paddiagram ziet er dan als volgt uit:

Diagram 1



Hierin worden p_{31} en p_{42} de stabiliteitscoëfficiënten genoemd en p_{32} en p_{41} de causaliteitscoëfficiënten.

De bij het Diagram 1 behorende structuurvergelijkingen zijn:

$$X_3 = p_{31} X_1 + p_{32} X_2 + e_3$$

$$X_4 = p_{41} X_1 + p_{42} X_2 + e_4$$

Uit de bovenstaande vergelijkingen kunnen we de volgende padcoëfficiëntenvergelijkingen afleiden, waaruit tevens de directe en indirecte effecten zonder meer zijn af te lezen:

$$r_{31} = p_{31} \text{ (dir. eff.)} + p_{32} r_{12} \text{ (indir. eff.)}$$

$$r_{32} = p_{31} r_{12} \text{ (indir. eff.)} + p_{32} \text{ (dir. eff.)}$$

$$r_{41} = p_{41} \text{ (dir. eff.)} + p_{42} r_{12} \text{ (indir. eff.)}$$

$$r_{42} = p_{41} r_{12} \text{ (indir. eff.)} + p_{42} \text{ (dir. eff.)}$$

Hieruit zijn de padcoëfficiënten op te lossen:

$$p_{31} = \frac{r_{13} - r_{12} r_{23}}{1 - r_{12}^2}$$

$$p_{32} = \frac{r_{23} - r_{12} r_{13}}{1 - r_{12}^2}$$

$$p_{41} = \frac{r_{14} - r_{12} r_{24}}{1 - r_{12}^2}$$

$$p_{42} = \frac{r_{24} - r_{12} r_{14}}{1 - r_{12}^2}$$

Voor ons voorbeeld: $p_{31} = .93$, $p_{32} = .02$, $p_{41} = .34$ en $p_{42} = .59$.

De correlaties r_{13} , r_{23} , r_{14} en r_{24} kunnen we dus opgebouwd zien uit de volgende directe en indirecte effecten:

$$r_{13} = .93 + .01$$

$$r_{23} = .49 + .02$$

$$r_{14} = .34 + .31$$

$$r_{24} = .18 + .59$$

We zien hier nog eens geïllustreerd dat de houding ten opzichte van de presidentskandidaat een te verwaarlozen effect heeft op stem-intentie. Resumerend stelt het gebruik van padcoëfficiënten ons in staat:

- samenhangen tussen variabelen te zien als opgebouwd uit een aantal directe en indirecte effecten;
- aan de hand van de vastgestelde directe effecten te bepalen welke variabele 'oorzaak' of tenminste meest dominante 'oorzaak' is;
- de relatieve grootte van de directe effecten vast te stellen (interpretatie van de padcoëfficiënten).

Als verder voordeel van padcoëfficiënten boven partiële correlatiecoëfficiënten valt aan te geven, dat padcoëfficiënten worden geschat vanuit een gespecificeerd model (structuurvergelijkingen), terwijl dit bij partiële correlatiecoëfficiënten niet het geval is.

Hier is immers alleen sprake van informatie omtrent het feit, dat een variabele al of niet méér onafhankelijk is dan een andere variabele.

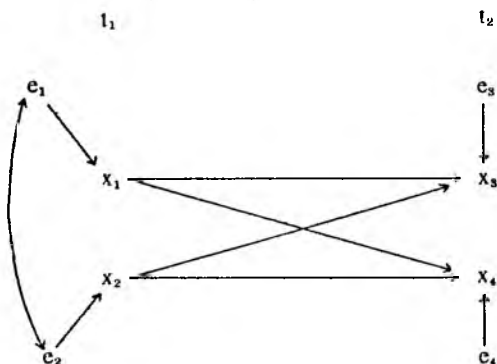
De veronderstellingen die we ten aanzien van de data maken bij toepassing van padanalyse, hanteren we ook hier: lineariteit, geen interactie, non-collineariteit, continue relaties, geen sprongsgewijze of drempel-niveau causaliteit, ongecorrleerd zijn van restvariabelen. Deze assumpties kunnen ten dele worden losgelaten. Zo kunnen we b.v. wel uitgaan van een correlatie tussen sommige restvariabelen. Dit stelt echter wel specifieke eisen aan het model (overidentificatie), aangezien anders de modelparameters niet meer rechtstreeks kunnen worden geschat. We komen hier in de volgende paragraaf op terug. Hanteren we padanalyse bij paneldata, dan krijgen we te maken met een aantal aanvullende veronderstellingen, die samenhangen met de veronderstelde vertraging³ van de effecten. Zo dienen we te assumeren, dat de meetperiode kleiner is dan de kleinste vertraging omdat anders niet meer is vol te houden dat de op tijdstip t_1 gemeten variabelen elkaar binnen dat tijdstip niet beïnvloeden. Verder dient het tijdsinterval tussen t_1 en t_2 groter te zijn dan de grootste vertraging, omdat anders de causale effecten nog niet de gelegenheid hebben gehad 'hun werk te doen'.

Willen beide assumpties redelijkerwijs tegelijkertijd opgaan, dan zal het duidelijk zijn dat we moeten veronderstellen dat alle vertragingen binnen het model ongeveer even groot zijn.

Gecorreleerde restvariabelen

Beschouwen we opnieuw het eerder gegeven paddiagram:

Diagram 2

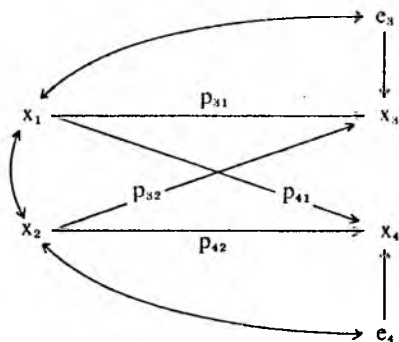


Bij de analyse hiervan zijn we tot nu toe steeds uitgegaan van de veronderstelling van ongecorreleerde restvariabelen (d.w.z. we veronderstellen $r_{e_1 e_3} = 0$ en $r_{e_2 e_4} = 0$). Deze veronderstelling lijkt echter weinig reëel onder het gezichtspunt dat e_1 en e_3 (en evenzo e_2 en e_4) dezelfde niet in het model opgenomen variabelen representeren op twee achtereenvolgende tijdstippen. Hetgeen bij enigszins stabiele variabelen tot een correlatie aanleiding zal geven.

In het navolgende zullen we nagaan wat het effect is van het niet gelijk nul veronderstellen van de correlaties $r_{e_1 e_3}$ en $r_{e_2 e_4}$ op de schatting van de overige modelparameters.

Daartoe gaan we uit van het onderstaande diagram, waarin e_1 en e_2 zijn weggelaten aangezien deze zonder meer verwisselbaar zijn met respectievelijk X_1 en X_2 . De correlaties r_{1e_3} en r_{2e_4} worden in het diagram aangegeven met gebogen pijlen evenals de correlatie tussen X_1 en X_2 .

Diagram 3



De structuurvergelijkingen m.b.t. het bovenstaande diagram zijn:

$$X_3 = p_{31} X_1 + p_{32} X_2 + p_{3e_3} e_3$$

$$X_4 = p_{41} X_1 + p_{42} X_2 + p_{4e_4} e_4$$

Vermenigvuldigen we de vergelijking van X_4 achtereenvolgens door met X_1 , X_2 , X_4 en e_4 :

$$r_{14} = p_{41} + p_{42} r_{12} + p_{4e_4} r_{1e_4}$$

Aangezien we veronderstellen, dat $r_{1e_4} = 0$ gaat de vergelijking over in:

$$r_{14} = p_{41} + p_{42} r_{12} \quad (1)$$

$$r_{24} = p_{41} r_{12} + p_{42} + p_{4e_4} r_{2e_4} \quad (2)$$

$$r_{44} = 1 = p_{41} r_{14} + p_{42} r_{24} + p_{4e_4} r_{4e_4} \quad (3)$$

$$r_{4e_4} = p_{41} r_{1e_4} + p_{42} r_{2e_4} + p_{4e_4} r_{e_4 e_4}$$

Aangezien $r_{1e_4} = 0$ en $r_{e_4 e_4} = 1$ gaat de laatste vergelijking over in:

$$r_{4e_4} = p_{42} r_{2e_4} + p_{4e_4} \quad (4)$$

Vermenigvuldigen we vergelijking (1) met r_{12} en trekken we vergelijking (2) hiervan af:

$$r_{12} r_{14} - r_{24} = p_{42} r_{12}^2 - p_{42} - p_{4e_4} r_{2e_4} \quad (5)$$

Vermenigvuldigen we nu vergelijking (1) met r_{14} en trekken we vergelijking (3) hiervan af:

$$r_{14}^2 - 1 = p_{42} r_{12} r_{14} - p_{42} r_{24} - p_{4e_4} r_{4e_4} \quad (6)$$

Uit vergelijking (5) volgt:

$$\begin{aligned} p_{4e_4} &= \frac{p_{42} r_{12}^2 - p_{42} + r_{24} - r_{12} r_{14}}{r_{2e_4}} \\ &= \frac{p_{42} (r_{12}^2 - 1) + (r_{24} - r_{12} r_{14})}{r_{2e_4}} \end{aligned} \quad (7)$$

Uit vergelijking (6) volgt:

$$\begin{aligned}
 p_{4e_4} &= \frac{p_{42} r_{12} r_{14} - p_{42} r_{24} + 1 - r_{14}^2}{r_{4e_4}} \\
 &= \frac{p_{42} (r_{12} r_{14} - r_{24}) + (1 - r_{14}^2)}{r_{4e_4}} \quad (8)
 \end{aligned}$$

Substitueren we r_{4e_4} uit vergelijking (8) in vergelijking (4), waarna we in de aldus ontstane nieuwe vergelijking p_{4e_4} uit vergelijking (7) substitueren. Dit resulteert in:

$$\begin{aligned}
 &\frac{p_{42}(r_{12}^2 - 1) + (r_{24} - r_{12}r_{14})}{r_{2e_4}} = \\
 &= \frac{p_{42} r_{2e_4} (r_{12} r_{14} - r_{24}) + r_{2e_4} (1 - r_{14}^2)}{p_{42} r_{2e_4}^2 + p_{42} (r_{12}^2 - 1) + (r_{24} - r_{12}r_{14})}
 \end{aligned}$$

Stellen we nu: $(r_{12} r_{14} - r_{24}) = a$

$$(1 - r_{14}^2) = b$$

$$(r_{12}^2 - 1) = c,$$

dan gaat bovenstaande vergelijking over in:

$$\frac{p_{42} c + a}{r_{2e_4}} = \frac{r_{2e_4} b - p_{42} r_{2e_4} a}{p_{42} r_{2e_4}^2 + p_{42} + a}$$

Deze laatste uitdrukking kunnen we schrijven als een kwadratische uitdrukking in p_{42} :

$$p_{42}^2 (r_{2e_4}^2 c + c^2) - p_{42} (2ac + 2r_{2e_4}^2 a) + (a^2 - r_{2e_4}^2 b) = 0$$

Via uitwerking van de wortelformule ter bepaling van oplossingen van de kwadratische vergelijking in p_{42} verkrijgen we voor p_{42} de volgende uitdrukking:

$$p_{42} = a/c \pm r_{2e_4} \sqrt{\frac{(a^2/c^2 + b/c)}{(r_{2e_4}^2 + c)}}$$

We kunnen nu voor p_{42} conditionele oplossingen bepalen voor achtereenvolgende waarden van r_{2e_4} , lopend van -1.00, -.90, via .00 tot +.90, +1.00.

Bij elk der gevonden waarden van p_{42} kunnen we uit vergelijking (1) de bijbehorende waarden van p_{41} berekenen.

Teneinde de veelheid van oplossingen voor p_{41} en p_{42} enigszins in te perken, zoeken we naar toetsingscriteria. Voldoen de oplossingen voor p_{41} en p_{42} niet aan deze criteria, dan zullen de oplossingen worden verworpen.

Een eerste toetsingscriterium wordt verkregen uit de vergelijkingen (4) en (7):

$$r_{4e_4} = \frac{p_{42} r_{2e_4}^2 + c p_{42} - a}{r_{2e_4}}$$

waarbij noodzakelijkerwijze dient te gelden: $-1.00 \leq r_{4e_4} \leq +1.00$.

Een tweede toetsingscriterium ontleen we aan het feit, dat het op grond van de op blz. 15 geformuleerde assumptie van het relatief constant zijn van de restvariabelen, niet plausibel is dat r_{2e_4} en r_{4e_4} ongelijk van teken zullen zijn.

Een derde toetsingscriterium kan worden verkregen uit vermenigvuldiging van de vergelijking voor X_4 met X_3 , waarbij verondersteld wordt, dat $r_{3e_4} = 0$. Dit resulteert in:

$$\hat{r}_{34} = p_{41} r_{13} + p_{42} r_{23}$$

Een afwijking tussen de gereproduceerde $\hat{r}_{34}$ en de empirische r_{34} groter dan .1 zullen we als niet acceptabel beschouwen.

Een vierde toetsingscriterium is de proportie verklaarde variantie in X_4 . Immers: $0 \leq R_{4.12}^2 \leq 1$.

Passen we bovenstaande werkwijze toe op de data van Lipset e.a., dan kunnen de resultaten in de tabel op blz. 20 worden weergegeven.

De acceptabele oplossingen voor p_{41} bewegen zich tussen de grenzen $-.68$ en $+1.35$; die voor p_{42} tussen -1.32 en $+2.51$.

Berekenen we voor elke combinatie van p_{41} en p_{42} de proportie verklaarde variantie in X_4 volgens de formule:

prop. verkl. var. = $p_{41} r_{14} + p_{42} r_{24}$, dan verkrijgen we het volgende resultaat:

Tabel 6

acceptabele oplossingen voor p_{41} en p_{42}		proportie verklaarde variantie in X_4
$-.68$	2.51	1.49 (onmogelijk)
1.35	-1.32	$-.12$ (onmogelijk)
$-.22$	1.64	1.12 (onmogelijk)
$.89$	$-.45$	$.23$
$.69$	$-.08$	$.39$
$.60$	$.10$	$.47$
$.53$	$.23$	$.52$
$.47$	$.34$	$.56$
$.42$	$.43$	$.60$
$.38$	$.51$	$.64$

Op grond van bovenstaande berekeningen krijgen we de volgende intervallen voor de waarden van p_{41} , p_{42} en r_{2e_4} :

$$.38 \leq p_{41} \leq .89$$

$$-.45 \leq p_{42} \leq .51$$

$$.10 \leq |r_{2e_4}| \leq .70$$

Op analoge wijze kunnen we een kwadratische uitdrukking voor p_{31} opstellen:

$$p_{31}^2 (cr_{1e_3}^2 + c^2) - p_{31} (2ar_{1e_3}^2 + 2ac) + br_{1e_3}^2 + a^2 = 0,$$

waarbij:

Tabel 5. Conditionele oplossingen voor p_{41} en p_{42} voor achtereenvolgende waarden voor r_{2e} lopend van -1.00 tot 1.00

r_{2e_4}	p_{42}	p_{41}	r_{4e_4}	teken r_{2e_4} en r_{4e_4} gelijk	$ \hat{r}_{34} - r_{34} $
-1.0	geen opl.	-	-	nvt	-
-.9	geen opl.	-	-	nvt	-
-.8	2.51	-.68	-.28	ja	.03
	-1.32	1.35	-.66	ja	.07
-.7	1.64	-.22	-.07	ja	.04
	-.45	.89	-.76	ja	.06
-.6	1.26	-.02	.05	nee	.04
	-.08	.69	-.76	ja	.06
-.5	1.08	.08	.17	nee	.05
	.10	.60	-.76	ja	.06
-.4	.95	.15	.26	nee	.05
	.23	.53	-.74	ja	.06
-.3	.85	.20	.36	nee	.05
	.34	.47	-.71	ja	.05
-.2	.76	.25	.44	nee	.05
	.43	.42	-.67	ja	.05
-.1	.67	.29	.51	nee	.05
	.51	.38	-.63	ja	.05
0.0	.59	.34	+.57	nvt	-
.1	.67	.29	-.52	nee	.05
	.51	.38	.63	ja	.05
.2	.67	.25	-.44	nee	.05
	.43	.42	.67	ja	.05
.3	.85	.20	-.36	nee	.05
	.34	.47	.71	ja	.05
.4	.95	.15	-.27	nee	.05
	.23	.53	.74	ja	.06
.5	1.08	.08	-.17	nee	.05
	.10	.60	.76	ja	.06
.6	1.26	-.02	-.05	nee	.04
	-.08	.69	.76	ja	.06
.7	1.64	-.22	.07	ja	.04
	-.45	.89	.76	ja	.06
.8	2.51	-.68	.28	ja	.03
	-1.32	1.35	.66	ja	.07
.9	geen opl.	-	-	nvt	-
1.0	geen opl.	-	-	nvt	-

$$a = (r_{12} r_{23} - r_{13})$$

$$b = (r_{23}^2 - 1)$$

$$c = (r_{12}^2 - 1)$$

Oplossingen voor p_{31} vinden we volgens:

$$p_{31} = a/c \pm r_{1e_3} \sqrt{\frac{(a^2/c^2 - b/c)}{(r_{1e_3}^2 + c)}}$$

Op grond hiervan kunnen we de conditionele oplossingen voor p_{31} en p_{32} voor achtereenvolgende waarden van $r_{1e_3} = .00, +.10, +.20, \dots, +1.00$ berekenen.

De resultaten hiervan worden in tabel 7 weergegeven.

De acceptabele oplossingen voor p_{31} bewegen zich tussen de grenzen $+ .72$ en $+1.52$; die voor p_{32} tussen $- .30$ en $+ .13$.

Berekenen we voor elke combinatie van p_{31} en p_{32} de proportie verklaarde variantie in X_3 volgens de formule:

$$\text{prop. verkl. var.} = p_{31} r_{13} + p_{32} r_{23}.$$

Tabel 8

acceptabele oplossingen voor p_{31} en p_{32}		proportie verklaarde variantie in X_3
1.52	-.30	1.28 (onmogelijk)
1.33	-.20	1.15 (onmogelijk)
1.23	-.14	1.08 (onmogelijk)
1.15	-.10	1.03 (onmogelijk)
.72	.13	.74
.78	.10	.78
.87	.05	.82
.88	.04	.85

Op grond van bovenstaande berekeningen krijgen we de volgende intervallen voor de waarden van p_{31} , p_{32} en r_{1e_3} :

Tabel 7. Conditionele oplossingen voor p_{31} en p_{32} voor achtereenvolgende waarden voor r_{1e} lopend van -1.00 tot 1.00

r_{1e_3}	p_{31}	p_{32}	r_{3e_3}	teken r_{1e_3} en r_{3e_3} gelijk	$ \hat{r}_{43} - r_{43} $
-1.0	geen opl.	-	-	nvt	-
-.9	geen opl.	-	-	nvt	-
-.8	2.07	-.59	-.62	ja	.22
	-.21	.62	-.86	ja	.44
-.7	1.52	-.30	-.46	ja	.09
	.35	.32	-.84	ja	.19
-.6	1.33	-.20	-.32	ja	.05
	.53	.23	-.80	ja	.15
-.5	1.23	-.14	-.18	ja	.02
	.64	.17	-.74	ja	.12
-.4	1.15	-.10	-.07	ja	.00
	.72	.13	-.67	ja	.10
-.3	1.08	-.06	.03	nee	.02
	.78	.10	-.60	ja	.09
-.2	.99	-.01	.02	nee	.04
	.87	.05	-.39	ja	.07
-.1	.98	-.01	.26	nee	.04
	.88	.04	-.45	ja	.07
0.0	.93	.02	+.34	nvt	-
.1	.98	-.01	-.26	nee	.04
	.88	.04	.45	ja	.07
.2	.99	-.01	-.02	nee	.04
	.87	.05	.39	ja	.07
.3	1.08	-.06	-.03	nee	.02
	.78	.10	.60	ja	.09
.4	1.15	-.10	.07	ja	.00
	.72	.13	.67	ja	.10
.5	1.23	-.14	.18	ja	.02
	.64	.17	.74	ja	.12
.6	1.33	-.20	.32	ja	.05
	.53	.23	.80	ja	.15
.7	1.52	-.30	.46	ja	.09
	.35	.32	.84	ja	.19
.8	2.07	-.59	.62	ja	.22
	-.21	.62	.86	ja	.44
.9	geen opl.	-	-	nvt	-
1.0	geen opl.	-	-	nvt	-

$$.72 \leq p_{31} \leq .88$$

$$.04 \leq p_{32} \leq .13$$

$$.10 \leq |r_{1e_3}| \leq .40$$

De intervallen voor de parameters p_{41} , p_{42} en r_{2e_4} bij het bovenstaande voegend geeft aanleiding tot de conclusie, dat, hoewel de intervallen vrij groot zijn, bij alle mogelijke waarden voor r_{1e_3} en r_{2e_4} de hypothese met betrekking tot de dominante oorzaak (t.w. stemintentie) overeind blijft.

Conclusie

In het algemeen gaan we bij causale analyse van paneldata uit van de veronderstelling van ongecorrleerde restvariabelen. Zijn deze restvariabelen in de empirie echter wel onderling gecorreleerd dan zijn de berekende waarden van de padcoëfficiënten 'biased'. Het loslaten van de veronderstelling van ongecorrleerde restvariabelen levert conditionele schattingen voor de parameters, waarbij we globaal een drietal situaties kunnen onderscheiden:

1. de schattingen voor de parameters liggen voor het merendeel binnen relatief grote intervallen, terwijl de hypothese omtrent de dominante oorzaak wordt weerlegd;
2. de schattingen voor de parameters liggen voor het merendeel binnen relatief grote intervallen en de hypothese omtrent de dominante oorzaak wordt niet weerlegd;
3. de schattingen van de parameters liggen binnen relatief kleine intervallen en de hypothese omtrent de dominante oorzaak wordt niet weerlegd.

De eerste situatie impliceert een verslechtering ten opzichte van de uitgangssituatie, de tweede situatie (gelijk aan die in het boven gegeven voorbeeld) doet de hypothese omtrent de dominante oorzaak sterker staan, de derde situatie geeft meer informatie dan in de uitgangssituatie het geval is.

- Blalock, H. M., *Causal inferences in non-experimental research*, Chapel Hill, 1964.
- Bohrnstedt, G. W., 1969, 'Observations on the measurement of change', in: E. F. Borgatta (Ed.), *Sociological methodology 1969*, San Francisco, 1969, pp. 113-134.
- Campbell, D. T., 'From description to experimentation: interpreting trends as quasi-experiments', in: Ch. W. Harris (Ed.), *Problems in measuring change*, Madison, 1963, pp. 212-242.
- Dessens, J., W. Jansen en P. G. Swanborn, 'Padanalyse: uitgangspunten en basisbegrippen', in: *Mens en maatschappij*, 49, 1974, pp. 157-184.
- Duncan, O. D., 'Some linear models for two wave, two variable panel analysis', in: *Psychological Bulletin*, 72, 1969, pp. 177-182.
- Duncan, O. D., 'Unmeasured variables in linear models for panel analysis', in: H. L. Costner (Ed.), *Sociological methodology 1972*, San Francisco, 1972, pp. 36-83.
- Gadourek, I. en H. R. Knol, 'Onderzoek naar sociaal culturele veranderingsprocessen in Nederland. Enkele methodologische dilemma's,' in: P. G. Swanborn (Ed.), *Methoden en mensen*, Rotterdam, 1974, pp. 91-121.
- Galtung, J., 'Diachronic correlation, process analysis and causal analysis', in: *Quality & Quantity*, 4, 1970, pp. 55-94.
- Heise, D. R., 'Causal inference from panel data' in: E. F. Borgatta and G. W. Bohrnstedt (Eds.), *Sociological methodology 1970*, San Francisco, 1970, pp. 3-28.
- Kenny, D. A., 'Cross-lagged and synchronous common factors in panel data', in: A. S. Goldberger and O. D. Duncan (Ed.), *Structural equation models in the social sciences*, New York, 1973, pp. 153-166.
- Lazarsfeld, P. F., 'Mutual effects of statistical variables', in: P. F. Lazarsfeld, A. K. Pasanella and M. Rosenberg (Eds.), *Continuities in the language of social research*, New York, 1972, pp. 388-398.
- Lipset, S. M., P. F. Lazarsfeld, A. H. Barton and J. Linz, 'The psychology of voting: an analysis of political behavior', in G. Lindzey (Ed.), *Handbook of social psychology*, Cambridge, 1954, pp. 1124-1175.
- Pelz, D. C. and F. M. Andrews, 'Detecting causal priorities in panel study data', in: *American Sociological Review*, 29, 1964, pp. 836-848.
- Rozelle, R. M. and D. T. Campbell, 'More plausible rival hypotheses in the cross-lagged correlation technique', in: *Psychological Bulletin*, 1969, 71, pp. 74-80.
- Yee, A. H. and N. L. Gage, 'Techniques for estimating the source and direction of causal influence in panel data', in: *Psychological Bulletin*, 70, 1968, pp. 115-126.

Noten

1. De turnover is het totaal aantal eenheden, dat tussen t_1 en t_2 op een variabele is veranderd.
2. Zie Dessens e.a., 1974.
3. Met het begrip vertraging wordt bedoeld het tijdsverloop tussen een verandering in een veroorzakende variabele en het effect van deze verandering op de veroorzaakte variabele.