KM23(1987)
pag 145-169

MODELLEN VOOR SERIËLE EFFECTEN IN BEOORDELINGEN

C.F. Maas, H.P.L. Seip en W.E. Saris

ABSTRACT

Met het toenemend gebruik van psychophysische meetmethoden in grootschalig enquete onderzoek probeert men de laatste jaren de kwaliteit van metingen in de sociale wetenschappen te verbeteren. Uit onderzoek is echter gebleken, dat bij het beoordelen van reeksen stimuli seriële effecten in de respons kunnen optreden, hetgeen consequenties heeft voor de schatting van de relatie tussen stimuli en responses.

In deze studie naar het optreden en modelleren van seriële effecten in relatie tot de grootte van de verschillen tussen de opeenvolgende stimuli en de gebruikte meetprocedures hopen we meer inzicht te verschaffen in dit probleem. Vervolgens komen we tot enige practische aanbevelingen, die zowel de kans op seriële afhankelijkheden tot een minimum beperken, als de vergelijkbaarheid van scores over respondenten vergroten.

Deze studie kon worden gerealiseerd dankzij subsidie van:

Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek,
projectnr. 430-213-005.

Universiteit van Amsterdam,
Vakgroep Methoden en Technieken van Politicologisch Onderzoek
Grimburgwal 10 geb. 5, 1012 GA AMSTERDAM
Tel: 020 - 5252089 (secr.)

Inleiding

De laatste jaren valt een toenemende belangstelling te constateren voor het gebruik van psychophysische meetprocedures in enquete onderzoek (Lodge 1981, Wegener 1982 en van Doorn et al. 1983). Belangrijke voordelen van dergelijke procedures t.o.v. de meer gebruikelijke metingen in categorieën zijn het hogere (log) interval meetniveau met de daarmee gepaard gaande winst aan informatie en de mogelijkheid tot herhaling van de meting door gebruik te maken van verschillende antwoordmodaliteiten. Dit stelt de onderzoeker in staat de betrouwbaarheid van de metingen te evalueren, zonder de validiteit nadelig te beïnvloeden (Saris, 1982). Door het gebruik van psychophysische meetprocedures wordt ook de mogelijkheid tot het benutten van krachtiger statistische analyse technieken vergroot.

Naast het veelal (te) lage meetniveau vormen de zwakke relaties tussen de gemeten variabelen een tweede belangrijk probleem bij statistische analyse in de sociale wetenschappen. Uit recente experimenten is gebleken dat een oorzaak hiervan mogelijk kan zijn gelegen in het feit dat niet iedereen op dezelfde manier zijn of haar mening omzet in een response, zodat variaties in responses niet voor eenieder op dezelfde manier corresponderen met de variatie in opinie t.a.v. de gepresenteerde stimulus (Saris et al., 1987). Intuïtief valt dit te begrijpen uit het feit dat sommige mensen overdrijven in het formuleren van hun mening, waar anderen juist relativeren. Dit verschijnsel is bekend onder de noemer variatie in individueel antwoordgedrag. Een belangrijk gevolg van dit verschijnsel is de onderschatting van de relaties tussen variabelen indien ze over de respondenten worden beschouwd. Door individueel antwoordgedrag te negeren wordt een niet onbelangrijke hoeveelheid ruis geïntroduceerd, hetgeen tot beduidend minder goede analyse resultaten kan leiden en zelfs tot foute conclusies.

Een en ander maakt het noodzakelijk in analyses rekening te gaan houden met de individuele manier waarop responses tot stand komen. Hiertoe dienen we te beschikken over kennis van de factoren die van invloed zijn op de relatie tussen de stimuli en de responses, zodat modellen kunnen worden geformuleerd die deze relaties goed kunnen beschrijven.

Uit experimenten met psychophysiche meetmethoden is bekend, dat de relatie tussen stimuli en responses kan worden beschreven door een power functie. Voor een overzicht

van de literatuur over dit onderwerp verwijzen we naar Stevens (1975).

$$R = a(S)^b \qquad \text{(vgl. 1)}$$

Hierin stelt R de response voor, a een schaalconstante , S de waarde van het stimulus en b de voor de gebruikte antwoordmodaliteit specifieke constante. Na logaritmische transformatie ontstaat aldus een lineaire relatie tussen de stimuluswaarden en de responses.

$$\ln(R) = \ln(a) + b \ln(S) \qquad \text{(vgl. 2)}$$

Om factoren die mogelijk een rol spelen in de relatie tussen stimuli en responses goed van elkaar te kunnen onderscheiden dient ook aandacht te worden besteed aan de in het experiment gebruikte meetprocedures (Saris et al., 1987). Enerzijds aan de manier waarop de stimuli aan respondenten worden aangeboden en anderszijds aan eigenschappen van de relatie tussen stimuli en responses die specifiek zijn voor de gebruikte antwoordmodaliteit.

Voor het schatten van de parameters van de modellen wordt doorgaans gebruik gemaakt van OLS. Hierbij wordt verondersteld dat de errortermen onafhankelijk zijn van elkaar. Indien respondenten bij het beoordelen van stimuli rekening houden met reeds eerder gegeven oordelen, ontstaan afhankelijkheden tussen de errortermen en wordt aan de voorwaarden van het model niet meer voldaan. Een dergelijke schending van de assumpties van het model kan resulteren in een bias in de schatting van de parameters (Johnston, 1972; p. 243 - 266, Theil, 1971; p. 256 - 257). Uit onderzoekingen van o.a. Green et al. (1977) en Cross (1973) is gebleken dat hiervan inderdaad sprake is.

In dit stuk zullen wij dit probleem beschouwen als uitdrukking van het feit dat aan de responses in deze situaties een ander proces ten grondslag ligt. In het experiment dat we voor de bestudering van de eventuele effecten van responses op de voorafgaande stimulus hebben opgezet, proberen we door vergelijking van een drietal eenvoudige modellen onder verschillende condities meer inzicht te krijgen in 1) waar doen de problemen zich voor, 2) welk model kan de verkregen responses goed en plausibel verklaren en 3) op welke wijze moet aan ongewenste effecten van eerder beoordeelde stimuli het hoofd worden geboden.

Om zo duidelijk mogelijke resultaten te krijgen, werd aan respondenten de eenvoudige taak voorgelegd om de lengte van verschillende lijnstukjes uit te drukken in een getal en wel zo dat men naarmate de lijn langer is ook een groter getal geeft. Uit de literatuur (Stevens, 1975) is bekend, dat de voor deze modaliteit specifieke constante de waarde 1 heeft. We kunnen daarom veronderstellen dat de relatie tussen stimuli en responses reeds lineair is. Als we geen rekening houden met seriële ahankelijkheden, dus onder de vooronderstelling dat de errortermen ongecorreleerd zijn met elkaar, kunnen we de relatie tussen stimuli en responses beschrijven met:

$$R = a(S) + c \qquad \text{(vgl. 3).}$$

Hierin is de constante c opgenomen voor correctie van het 0-punt. Dit basismodel gaan we vervolgens vergelijken met een tweetal modellen waarin, middels opname van termen in de vergelijkingen voor de response op de voorafgaande stimulus, wél rekening is gehouden met seriële afhankelijkheden in de responses.

Alternatieve response modellen

In het algemeen wordt bij de beschrijving van de relatie tussen stimuli en responses in psychopysische experimenten geen rekening gehouden met seriële afhankelijkheden in de oordelen. Indien echter dergelijke afhankelijkheden toch aanwezig zijn, hebben we te maken met een misspecificatie van het model ten gevolge waarvan de errortermen gecorreleerd zullen zijn hetgeen, zoals eerder reeds opgemerkt, schending van de modelassumpties betekent en zal resulteren in een bias in de schatting van de parameters. Als alternatieven voor het eenvoudige lineaire model (vgl.3) worden in deze studie een tweetal verschillende modellen bekeken waarin rekening wordt gehouden met het optreden van seriële afhankelijkheid in de response.

In de literatuur wordt dit probleem meestal opgelost door eenvoudig aan de oorspronkelijke vergelijking een term toe te voegen die R_{t-1} bevat. Uitgaande van vgl.3 krijgen we zo:

$$R_t = aS_t + bR_{t-1} + c + e_t \qquad \text{(vgl. 4)}$$

(In deze vergelijking is verondersteld dat de errortermen ongecorreleerd zijn met elkaar en met de exogene variabelen)

Hierdoor wordt voor correlaties tussen de errortermen gecorrigeerd, voor zover deze het gevolg zijn van een effect van de response op de voorafgaande stimulus. Ook in de studie van Cross werd van een dergelijk model gebruik gemaakt, zij het dat de variabelen daar in log-scores gegeven waren. Beoordeling van een aangeboden stimulus ontstaat in dit model als het ware als een gewogen gemiddelde van de voorliggende stimulus en de response op de voorafgaande stimulus. Het valt echter niet aan te nemen dat respondenten ook inderedaad een dergelijke calculatie maken, daarom moeten we dit model beschouwen als een weergave van een proces waarin de respondent, zonder het zich bewust te zijn, zijn oordelen laat beïnvloeden door eerder gegeven antwoorden. De vraag waaróm respondenten juist op deze manier (onwillekeurig) rekening zouden houden met eerder gegeven oordelen blijft onbeantwoord. Om meer inzicht te krijgen in wat er zich afspeelt, zullen de resultaten van dit model worden vergeleken met een model waarin afhankelijkheden tussen oordelen van opvolgende stimuli op een andere manier worden verklaard. Voor schatting van de parameters van dit model dient het experiment de drietallen (R_t, S_t en R_{t-1}) op te leveren.

Een meer voor de hand liggende manier om naar het probleem van afhankelijkheden tussen responses op opvolgende stimuli te kijken is dat een beoordelingstaak in de eerste plaats een *taak* is, die aan de respondenten wordt opgelegd. Het is zeker niet uitgesloten dat men probeert deze taak te vervullen met een *minimum* aan investering van tijd en moeite. Een eenvoudige reden voor het ontstaan van seriële afhankelijkheden zou daarom kunnen zijn dat respondenten proberen de beoordelingstaak te simplificeren door juist uit te gaan van de response op de voorafgaande stimulus en dit oordeel vervolgens corrigeren voor het verschil tussen de stimuli S_t en S_{t-1}. Vooral daar waar stimuli veel op elkaar lijken, dus bij kleine verschillen tussen de opvolgende stimuli, lijkt deze gedachtengang plausibel. Daar vormt R_{t-1} immers al een goede benadering voor R_t en is nog slechts een kleine correctie vereist. Volgen we deze redenering, dan ontstaat een model, dat in een vergelijking als volgt kan worden weergeven:

$$R_t = R_{t-1} + b(S_t - S_{t-1}) + e_t \qquad \text{(vgl. 5)}$$

Onder de gebruikelijke model veronderstellingen.

Dit is het derde model dat in deze studie zal worden gebruikt. Voor het schatten van de parameters dient het experiment de viertallen (R_t,S_t,S_{t-1},R_{t-1}) op te leveren.

We zijn er bij model 2 (vgl.4) impliciet van uit gegaan, dat de beïnvloeding van de response door de response op de voorafgaande stimulus constant is voor elk stimuluspaar, ongeacht de waarde van de voorafgaande response of de plaats die deze inneemt in de reeks. Model 3 (vgl.5) is opgesteld met het idee dat dit niet zo zou zijn en dat het juiste model bij kleine verschillen anders is als bij grote verschillen. Dit is in overeenstemming met de bevindingen van Green et al. (1977, p.455) waar zij concluderen:

> "The correlation between responses for each stimulus pair and averaged over constant stimulus differences (...) is (...) very large for small differences and about zero for large ones".

Om te kunnen controleren of ook inderdaad verschillende response modellen optreden, werd de aangeboden reeks van stimuli zodanig geconstrueerd, dat deze achteraf kon worden opgesplitst in twee subreeksen waarin respectievelijk slechts grote- dan wel kleine verschillen van constante grootte tussen opvolgende stimuli aanwezig waren.

Bij de hiervoor besproken modellen is geen rekening gehouden met 2^e-orde effecten. Hiermee wordt bedoeld dat de response op tijdstip t wordt beinvloed door de response op tijdstip $t\text{-}2$. In de studie van Green et al. bleek de invloed van deze 2^e-orde effecten echter zeer gering te zijn (p.453). We hebben ons daarom beperkt tot het bekijken van modellen met 1^e-orde effecten.

Opzet van het experiment

Aan een tiental respondenten werd gevraagd beoordelingen te geven van de lengtes van een reeks lijnstukken. De oordelen moesten worden uitgedrukt in een getal, dat correspondeerde met de lengte van het betreffende lijnstuk. Voor het afnemen van de interviews werd gebruik gemaakt van een microcomputer, alsmede van het programma INTERV (de Pijper et al., 1986). De respondenten kregen voor dit experiment een reeks van 30 lijnstukken te beoordelen. Nadat zij deze taak hadden vervuld moesten enige reeksen met andere stimuli worden beoordeeld, waarna de eerste beoordelingstaak met exact dezelfde stimuli in exact dezelfde volgorde werd herhaald. Nu was in de vraagstelling echter een standaardlijn met een vaste modulus opgenomen. In het vervolg zullen we de eerste reeks (zonder standaard en modulus) aanduiden met *Blok 1* en de herhaling (met standaard en modulus) met *Blok 7*. Werd de respondent in zijn beantwoording in blok 1

volledig vrij gelaten, in blok 7 moesten de oordelen worden gerelateerd aan een lijn van standaard lengte en een vaste response waarde.

In blok 1 werd de taak als volgt geformuleerd:

> "Op dit scherm zal telkens een lijnstuk verschijnen. We vragen
> U de lengte van het lijnstuk weer te geven in een getal.
> Is de lijn *lang* , geef dan een *groot* getal.
> Is de lijn *kort* , tik dan een *klein* getal in.
>
> Ken aan de eerste lijn een willekeurig getal toe."

En vervolgens telkens:

> "Ken een getal toe aan de volgende lijn."

In blok 7 luidde de vraagstelling:

> "hieronder ziet U een lijn.
> _____ (lijnstuk van lengte 40)
>
> Deze lijn noemen we standaardlijn en geven we het getal 500.
>
> Vindt U een lijn *langer* dan de standaardlijn, geef dan een *groter* getal.
> Vind U een lijn *korter* dan de standaardlijn, tik dan een *kleiner* getal in.
>
> Ken een getal toe aan de volgende lijn."

Achtereenvolgens moesten de ondervraagden lijnstukken beoordelen met respectievelijk lengtes van 12, 42, 46, 16, 20, 50, 54, 24, 28, 58, 62, 32, 36, 51, 59, 70, 47, 36, 32, 62, 58, 28, 24, 54, 50, 20, 16, 46, 42, 12.

De volgorde van de stimuli is niet willekeurig gekozen. In de eerste plaats valt te zien dat de eerste 13 stimuli aan het eind van de reeks in omgekeerde volgorde terugkeren, gescheiden door een viertal andere stimuli. In de tweede plaats nemen de verschillen tussen

opvolgende stimuli, in deze reeks van 13, (absoluut) slechts twee waarden aan, te weten 30 en 4. In de gehele reeks komen deze waarden 12 keer voor, 6 maal met positief- en 6 maal met negatief teken. Dit om een eventueel effect, dat zou kunnen worden veroorzaakt door de richting van de stimulusverschillen, bij voorbaat te neutraliseren. Uit de totale reeks van 30 stimuli kunnen we twee nieuwe reeksen construeren met constante verschillen tussen de opvolgende stimuli. De reeks met de grote verschillen (lengte 30), heeft telkens betrekking op de 2^e, 4^e, 6^e, 8^e, 10^e, 12^e, 20^e, 22^e, 24^e, 26^e, 28^e en 30^e stimulus voor tijdstip t en op de 1^e, 3^e, 5^e, 7^e, 9^e, 11^e, 19^e, 21^e, 23^e, 25^e, 27^e en 29^e stimulus voor het tijdstip $t\text{-}1$. Evenzo geldt voor de reeks met de kleine verschillen (lengte 4), dat deze betrekking hebben op de 3^e, 5^e, 7^e, 9^e, 11^e,13^e,19^e, 21^e, 23^e, 25^e, 27^e en 29^e stimulus voor tijdstip t en op de 2^e, 4^e, 6^e, 8^e, 10^e, 12^e, 18^e, 20^e, 22^e, 24^e, 26^e en 28^e stimulus voor het tijdstip $t\text{-}1$. Door de oordelen te beschouwen als response op de variabele "lijnlengte" beschikken we zo per persoon over 29 waarnemingen in het geval we de gehele reeks beschouwen en over telkens 12 waarnemingspunten indien we de reeksen voor grote- of kleine verschillen beschouwen.

Ongeveer een week na afname van het interview werd het experiment met dezelfde respondenten herhaald. Na screening van de data voor eventueel door de respondent gemaakte typfouten, beschikten we over twee vergelijkbare datasets, waarin zich gegevens bevinden over elk van de drie te analyseren reeksen voor de variabelen R_t, R_{t-1}, S_t en S_{t-1} per persoon en per blok.

Voor de beoordelingstaak werden een 10-tal studenten bereid gevonden. Omdat het aantal stimuli en niet het aantal respondenten bepalend is voor het aantal waarnemingen leek dit aantal voldoende voor het verkrijgen van een eerste indruk. Vanwege de eenvoud van de gestelde taak meenden wij te mogen veronderstellen dat de resultaten onafhankelijk zouden zijn van variabelen als opleiding etc., zodat afgezien kon worden van het trekken van een "echte" steekproef en de hogere kosten die dit met zich mee gebracht zou hebben.

Methode

Zoals uit het voorafgaande duidelijk zal zijn geworden, gaan we er van uit dat de variatie in de responses wordt veroorzaakt door;

- variatie in de lengte van de ter beoordeling gegeven lijnen

- individuele variatie in de manier van beantwoorden door de respondenten
- verschillende manieren van het aanbieden van de stimuli en
- seriële afhankelijkheden

Om te bepalen wélk model in wélke situatie het best de relatie tussen stimuli en response voor een specifieke respondent weergeeft, dienen we te beschikken over een computer-programma dat in staat is individuele data te analyseren. Voor de analyse is gebruik gemaakt van het programma Percase (de Pijper, 1982). Dit programma herstructureert de datamatrix zodanig, dat per case een nieuwe matrix wordt gevormd waarop analyses kunnen worden uitgevoerd. De resultaten van de berekeningen kunnen indien gewenst vervolgens weer aan een SPSS-file worden toegevoegd. In Percase staan o.a. ter beschikking; scattergrams, enkelvoudige regressie, log-transformatie voor variabelen en Lisrel 4, dat vanwege de mogelijkheid van een toets op de fit van het model goed toegesneden was op de behoefte van dit onderzoek.

De Lisrel-methode heeft verschillende voordelen boven het gebruik van regressie. In de eerste plaats wordt het mogelijk de nadruk in de analyse te verleggen van de verklaring van de variantie in R_t naar de verklaring van de covariantiestruktuur tussen de variabelen R_t, S_t, S_{t-1} en R_{t-1}. Dit is van belang, omdat de verschillende alternatieve modellen verschillende restricties opleggen aan de covarianties tussen de variabelen. Deze restricties zorgen voor een toesingsmogelijkheid van de fit van het model op de data. De toetsingsgrootheid is bij benadering χ^2 verdeeld, met het aantal vrijheidsgraden gelijk aan het aantal restricties op het volledig verzadigde model. Voor een meer uitgebreide behandeling verwijzen we naar Jöreskog & Sörbom (1978). Ook kunnen we, middels de test-statistic T, geneste modellen, d.w.z. modellen die uit een eenvoudiger model ontstaan door vrijmaking van een of meer parameters vergelijken. Een test-statistic op de uitbreiding van het model is gegeven door $T_u = T_a - T_b$. Ook deze grootheid is bij benadering χ^2 verdeeld met het aantal vrijheidsgraden $df = df_a - df_b$, waarbij a het meest eenvoudige model is. Tenslotte kunnen we multi-sample toetsen uitvoeren, om te controleren of een bepaald model in een bepaalde situatie over alle respondenten gezien als passend kan worden beschouwd. Hierbij is de toetsings grootheid $T_{tot} = T_1 + T_2 + \ldots\ldots\ldots + T_{10}$, bij benadering χ^2 verdeeld met $df = df_1 + df_2 + \ldots\ldots + df_{10}$. De test-statistic zal in het vervolg worden aangeduid met T.

Voorts kan Lisrel, onder bepaalde condities, maximum likelihood schatters voor de parameters van het model berekenen. Maximum likelihood schatters hebben het voordeel dat zij schaal-invariant zijn, d.w.z. dat de schattingen voor gestandariseerde- en ongestandariseerde oplossingen aan elkaar gelijk zijn op een schaaltransformatie na. Dit in tegenstelling tot Unweighted Least Squares (ULS) schatters, die deze eigenschap niet bezitten.

Omdat we in deze studie worden geconfronteerd met een zeer klein aantal waarnemingen verdienen de onderwerpen toetsing en schatting nog een korte bespreking. Met name dienen we ons af te vragen of;

a we de test statistic goed kunnen interpreteren, m.a.w. is deze nog χ^2 verdeeld
b misspecificatie van een model middels de test statistic kan worden gedetecteerd en
c de schattingen van de parameters en hun varianties voor ons doel voldoende nauwkeurig zijn.

Ad a: In de studie van Boomsma (1983) wordt de met het juiste model verkregen verdeling van de empirisch gevonden waarden van de test statistic vergeleken met de centrale χ^2 verdeling. Op p.119 schrijft hij dat voor $n \leq 50$:

> "*Depending on the model under study* the right tail of its sampling distribution may be either too heavy or too light" (cursivering van ons).

Op basis van de studie van Satorra en Saris (1982, p.174) mogen wij aannemen dat deze afwijkingen het gevolg zijn van de gekozen waarden van de parameters in de modellen die door Boomsma in zijn simulatie studie zijn onderzocht. Uit de studie van Satorra en Saris blijkt dat daar waar de verklaarde variantie in de modellen hoog is, zelfs voor steekproeven van 25 cases de verdeling van de test-statistic bij benadering χ^2 is. Omdat we in deze studie ook te maken hebben met hoge verklaarde varianties ($r^2 > .90$), menen we toch de χ^2 verdeling te kunnen gebruiken bij de toetsing.

Ad b: Het onderscheidend vermogen (power) van de toets blijkt eveneens sterk afhankelijk van de waarde van de parameters in het model. Naarmate de effecten sterker zijn en dus de verklaarde variantie toeneemt wordt ook de test statistic gevoeliger voor specificatie fouten in het betreffende (deel van) het model (Saris, Satorra & Sörbom, 1987). Om de reeds

onder a genoemde reden kunnen we veronderstellen dat misspecificaties in het model ook worden gedetecteerd.

Ad c: In deze studie zijn wij vooral geïnteresseerd in het al of niet bestaan van een effect. Daarom moeten we kunnen vertrouwen op de schattingen van de standaardfouten van de parameters. Boomsma merkt hierover op (p.117) :

> "...when correlation matrices are analized ... a severe overestimation of these variances can be expected for small sample sizes "

Om deze reden zullen we te snel geneigd zijn om insignificante effecten te vinden, in het bijzonder kan dit een rol spelen bij de evaluatie van het effect γ_{13} van de response op de voorafgaande stimulus in model 2. Hiermee zullen we dan ook rekening moeten houden in de analyses.

Hiërarchie en keuze-criteria

Voordat we de resultaten van de analyses zullen bespreken, zullen we eerst de te onderzoeken modellen en hun specifieke restricties aan een nader onderzoek onderwerpen en criteria opstellen om tot een rangordening te komen.

De modellen die corresponderen met de vergelijkingen 3,4, en 5 kunnen worden beschouwd als bijzondere gevallen van het onderstaande volledig verzadigde algemene model;

$$R_t = \gamma_{11}S_t + \gamma_{12}S_{t-1} + \gamma_{13}R_{t-1} + \zeta \qquad \text{(vgl.6)}$$

waarbij df = 0 en de variabelen in deviatiescores.

Voor model 1 (vgl. 3) geldt: $\gamma_{12} = \gamma_{13} = 0.$
Voor model 2 (vgl. 4) geldt: $\gamma_{12} = 0.$
Voor model 3 (vgl. 5) geldt: $\gamma_{12} = -\gamma_{11}$ en $\gamma_{13} = 1$

Het aantal vrijheidsgraden voor de modellen 1, 2 en 3 bedraagt respectievelijk 2, 1 en 2, hetgeen ook onmiddelijk is te zien aan het aantal restricties dat aan het algemene model moet worden opgelegd om tot deze modellen te komen. Ook is te zien dat model 1 kan

worden verkregen uit model 2 door de extra restrictie $\gamma_{13} = 0$ op te leggen. Bij de modellen 1 en 2 hebben we dan ook te maken met geneste modellen, zodat een toets mogelijk is op de uitbreiding van model 1 met de term $\gamma_{13}R_{t-1}$. Voor een keuze tussen de modellen 1 en 2 beschikken we aldus over een kriterium.

Uit verdere inspectie van de restricties wordt duidelijk dat de modellen 2 en 3 onderling niet op deze wijze vergelijkbaar zijn vanwege de *verschillende* restricties die in deze modellen aan γ_{12} worden opgelegd. Een criterium om te kiezen tussen deze modellen bestaat daarom in deze zin niet. We zijn in dit geval aangewezen op vergelijking van de verklaarde variantie, waarbij moet worden bedacht dat model 3 één vrijheidsgraad meer bezit.

Iets dergelijks geldt ook voor vergelijking van de modellen 1 en 3, vanwege de ook hier weer *verschillende* restricties die worden opgelegd aan de parameters γ_{12} en γ_{13}. Een echte test op het verschil tussen deze modellen is daarom ook in dit geval niet mogelijk. Vergelijking zal weer moeten plaatsvinden op basis van de verklaarde variantie. In twijfelgevallen geven we de voorkeur aan model 1.

Om een indruk te krijgen welke waarde moet worden gehecht aan de met de verschillende modellen verkregen resultaten, moet ook worden bekeken of de gevonden modellen in het tweede interview opnieuw als passend kunnen worden beschouwd. Indien dit niet het geval is bestaat het risico dat de resultaten berusten op toevallige fluctuaties in de responses. Replicatie van eenzelfde model in het tweede interview kunnen we opvatten als aanwijzing voor de juistheid van het gepostuleerde model.

Resultaten 1e interview

In model 1 worden de covarianties tussen R_t en R_{t-1} en tussen R_t en S_{t-1} bepaald door de covarianties tussen de exogenen onderling en door het directe effect γ_{11} van de stimulus op de response. Van de schatting van model 1 is de gestandariseerde oplossing weergegeven in tabel 1.

Kijkend naar de fit van dit model, valt op dat het vooral bij kleine verschillen tussen de opvolgende stimuli in blok 1 niet voldoet. In blok 7, wat alleen van blok 1 verschilt vanwege de in de vraagstelling opgenomen standaard, blijkt model 1 veel beter te passen.

Tabel 1: Resultaten model 1, df=2 (gestandariseerde oplossing)

	Blok 1				Blok 7			
resp.	γ_{11}	r^2	T	p	γ_{11}	r^2	T	p
alle versch. n=29								
1	.97	.95	14.61	.00*	.99	.99	.89	.64
2	.98	.97	1.76	.42	1.00	1.00	.41	.81
3	.96	.93	6.08	.05*	.99	.98	.52	.77
4	.95	.90	10.29	.01*	.99	.99	.61	.74
5	.95	.90	1.29	.52	1.00	.99	6.61	.04*
6	.96	.92	8.26	.02*	.98	.96	.56	.75
7	.92	.84	8.12	.02*	.99	.99	.30	.86
8	.96	.92	1.74	.48	.97	.95	.34	.84
9	.99	.98	5.52	.06	1.00	.99	3.05	.22
10	.99	.97	.99	.61	1.00	1.00	.40	.82
T_{tot} df=20			58.66*				13.69	
grote versch. n=12								
1	.96	.93	4.23	.12	.99	.98	4.34	.11
2	.99	.98	4.63	.10	1.00	1.00	.35	.84
3	.97	.93	1.60	.45	.99	.99	3.84	.15
4	.97	.93	7.27	.03*	.99	.99	2.88	.24
5	.96	.92	1.01	.60	1.00	1.00	15.79	.00*
6	.94	.89	5.93	.05	.98	.97	1.27	.53
7	.87	.75	1.65	.44	1.00	.99	.65	.72
8	.97	.93	7.99	.02*	.96	.93	10.70	.01*
9	.99	.98	3.37	.19	1.00	1.00	3.23	.20
10	.98	.97	5.16	.08	1.00	1.00	2.28	.32
T_{tot} df=20			42.84*				45.33*	
kleine versch. n=12								
1	.99	.98	6.08	.05*	1.00	.99	3.44	.18
2	.99	.98	14.99	.00*	1.00	1.00	2.08	.35
3	.95	.89	11.06	.00*	.99	.98	6.47	.04*
4	.93	.86	13.16	.00*	1.00	.99	.93	.63
5	.94	.89	25.54	.00*	1.00	.99	2.68	.26
6	.99	.97	9.33	.01*	.98	.96	10.14	.01*
7	.94	.88	11.85	.01*	.99	.99	.10	.95
8	.93	.87	30.33	.00*	.98	.95	5.82	.05
9	.99	.98	11.13	.01*	1.00	.99	1.09	.58
10	.99	.98	9.15	.01*	1.00	.99	7.94	.02*
T_{tot} df=20			142.62*				40.69*	

(kritieke waarde bij $\alpha = 5\%$ χ^2_{20} : 31.41 χ^2_2 : 5.99 / * = T is significant)

Omdat we juist in de reeks met de kleine stimulus verschillen de grootste effecten mogen verwachten van de response op het voorafgaande stimulus, kunnen we hierin een aanwijzing zien dat seriële effecten inderdaad lijken op te treden. Het aanbieden van een standaard in de vraagstelling lijkt deze effecten echter belangrijk te verminderen. Evaluatie van de andere twee modellen, waarin wél met effecten van het voorafgaande antwoord rekening is gehouden zal hierover meer inzicht moeten verschaffen.

Opvallend is ook de gunstige invloed van het aanbieden van een standaard op de verklaarde variantie, die ook in blok 1 al zeer hoog te noemen was. In blok 7 blijkt de proportie verklaarde variantie nog met gemiddeld 4,5% toe te nemen. In relatieve termen is dit 64% van de in eerste instantie verkregen restvariantie. Over respondenten gezien, voldoet model 1 alleen voor de complete reeks met alle stimulus verschillen, indien een standaard in de vraagstelling was aangeboden. Uit deze resultaten mogen we concluderen dat door het aanbieden van een standaard in de vraagstelling de resultaten in verschillende opzichten kunnen worden verbeterd. Verdere analyse zal moeten uitwijzen of met de modellen 2 en 3 een betere fit kan worden verkregen dan met model 1. De resultaten voor model 2 staan weergegeven in tabel 2.

In model 2 is de verklaring van de samenhang tussen R_t en R_{t-1} en tussen R_t en S_{t-1} niet meer uitsluitend afhankelijk van de schatting van γ_{11}, maar ook van γ_{13}. Omdat de samenhang tussen de response variabelen bij de kleine stimulusverschillen het sterkst is, zal het effect van de response op het voorafgaande stimulus op het antwoord hier ook het grootst zijn, zoals duidelijk valt te zien aan de resultaten van blok 1.

In blok 7 ligt de situatie anders. Op basis van de resultaten van model 1 liet zich al vermoeden dat hier in sterk verminderde mate sprake zou zijn van seriële effecten. Model 2 paste in blok 7 in vrijwel alle gevallen goed op de data. Het werd slechts twee keer verworpen voor de grote verschillen en drie keer voor de kleine verschillen. De schattingen voor het effect γ_{13} zijn in dit blok overal laag (m.u.v. resp 6, kleine versch.) en in vrijwel alle gevallen ook niet significant verschillend van 0. Kennelijk kon de samenhang tussen R_t en respectievelijk S_{t-1} en R_{t-1} hier al voldoende worden verklaard via het effect γ_{11}. Bij het insignificant zijn van deze effecten zal echter ook het geringe aantal waarnemingen en de hoge verklaarde variantie die reeds met model 1 kon worden bereikt een rol hebben gespeeld. Significante effecten treffen we vooral aan bij de kleine verschillen in blok 1. Toch moest het model hier bij zes van de tien respondenten verworpen worden.

Tabel 2: Resultaten model 2, df=1 (gestandariseerde oplossing)

	Blok 1					Blok 7				
resp.	γ_{11}	γ_{13}	r^2	T	p	γ_{11}	γ_{13}	r^2	T	p
alle versch. n=29										
1	.95	.14	.97	2.19	.14	.99	.01^x	.99	.82	.37
2	.98	.04^x	.97	.62	.43	.99	.01^x	1.00	.12	.73
3	.95	.05^x	.93	5.10	.02*	.99	.02^x	.98	.03	.86
4	.94	.15	.92	4.10	.04*	.99	.02^x	.99	.01	.94
5	.94	.05^x	.90	.62	.43	.99	.04	.99	.06	.81
6	.93	.15	.95	.08	.78	.98	.02^x	.96	.39	.53
7	.88	.18	.87	2.78	.10	.99	.01^x	.99	.03	.86
8	.95	.04^x	.92	.86	.35	.97	.03^x	.95	.00	.98
9	.98	.03^x	.98	4.68	.03*	1.00	.00^x	.99	3.04	.08
10	.98	.03^x	.97	.08	.78	1.00	-.01^x	1.00	.01	.92
T_{tot} df=10				21.11*					4.51	
grote versch. n=12										
1	1.09	.20	.95	.46	.50	.97	-.03^x	.99	3.88	.05*
2	1.06	.12	.99	.30	.58	1.00	.00^x	1.00	.11	.74
3	1.02	.11^x	.94	.08	.78	1.04	.07^x	.99	.63	.43
4	1.12	.24	.97	.02	.88	.95	-.06^x	.99	1.08	.30
5	1.01	.08^x	.93	.39	.53	1.01	.08^x	1.00	.02	.88
6	1.11	.26	.93	1.18	.28	1.01	.04^x	.97	.91	.34
7	.95	.15^x	.77	.93	.34	1.01	.03^x	.99	.02	.89
8	.98	.03^x	.93	7.90	.01*	1.01	.08^x	.93	10.08	.00*
9	1.00	.02^x	.98	3.25	.07	1.00	.01^x	1.00	3.17	.08
10	1.01	.04^x	.97	4.89	.03*	1.00	.00^x	1.00	2.27	.13
T_{tot} df=10				19.40*					22.17*	
kleine versch. n=12										
1	.65	.36	.99	.31	.58	.84	.17^x	.99	1.50	.22
2	.68	.33	.99	4.52	.03*	.91	.09^x	1.00	.06	.81
3	.76	.20^x	.90	10.51	.00*	.97	.02^x	.98	6.45	.01*
4	.36	.64	.95	1.94	.16	1.06	-.07^x	.99	.60	.44
5	.37	.64	.98	8.55	.00*	.96	.03^x	.99	2.55	.11
6	.77	.24	.99	2.68	.10	.56	.44	.98	3.53	.06
7	.68	.37	.96	1.96	.16	1.02	-.03^x	.99	.04	.85
8	.41	.59	.94	21.81	.00*	.89	.10^x	.95	5.56	.02*
9	1.45	-.47	.98	7.70	.01*	.97	.03^x	.99	.99	.32
10	.73	.27	.98	5.75	.02*	.90	.10^x	1.00	6.50	.01*
T_{tot} df=20				65.73*					27.78*	

(k.w. bij $\alpha = 5\%$ χ^2_{10}: 18.31 χ^2_1: 3.84 / * = T is significant / $^x = \gamma_{13}$ niet significant)

In de reeks met grote verschillen en de reeks met alle verschillen was het beter gesteld met de fit van het model, maar hebben we om de boven al genoemde redenen te maken met insignificante effecten. De verklaarde variantie ging met model 2 t.o.v. model 1 alleen in de reeks met de kleine verschillen noemenswaardig omhoog, gemiddeld met bijna 4%. Over alle respondenten gerekend, past model 2, net als model 1, alleen op de complete reeks, indien er een standaard in de vraagstelling was opgenomen.

We konden nu nagaan, middels de toets op het verschil in T, of model 2 significante verbetering in fit te zien gaf t.o.v. model 1. De resultaten zijn weergegeven in tabel 3.

Tabel 3: $T = T_{mod1} - T_{mod2}$ met $df = 1$

	Blok 1			Blok 7		
resp.	groot	klein	alle	groot	klein	alle
1	3.77	5.77*	12.43*	.46	1.94	.07
2	4.33*	10.46*	1.14	.24	2.02	.29
3	1.52	.55	.98	3.21	.02	.49
4	7.25*	11.22*	6.19*	1.80	.33	.60
5	.62	16.99*	.67	15.77*	.13	6.55*
6	4.76*	6.77*	8.19*	.36	6.62*	.17
7	.72	9.89*	5.34*	.63	.06	.27
8	.09	8.52*	.88	.62	.26	.34
9	.12	3.44	.84	.06	.10	.01
10	.27	3.40	.91	.01	1.44	.39
T_{tot}	23.45*	77.01*	37.57*	23.16*	12.92	9.18

(kritieke waarde bij $\alpha = 5\%$ χ^2_{10} : 18.31 χ^2_1 : 3.84 / * = T is significant)

De verbetering die met model 2 kan worden bereikt is verreweg het grootst en treedt het meest in significante proporties op, bij de kleine verschillen in blok 1. Ten overvloede merken we op dat waar een significante verbetering in fit optreedt, model 2 toch niet in alle gevallen mag worden aanvaard. Voor blok 1 geeft model 2 vaak een significante verbetering t.o.v. model 1 te zien, ook over respondenten, met name in de reeks met kleine verschillen. We kunnen het echter slechts in individuele gevallen als alternatief beschouwen. Niet alleen omdat model 2 niet overal op de data past, maar ook omdat in niet

alle gevallen het effect γ_{13} significant is.

Voor blok 7 geldt dat model 2 vrijwel nergens significante verbetering te zien gaf t.o.v. model 1, ook niet over respondenten gezien. Enige uitzondering vormde wat dit laatste betreft de reeks met de grote verschillen. Uit de tabel wordt echter duidelijk dat dit vrijwel geheel valt toe te schrijven aan respondent 5.

Combineren we de gegevens en stellen we de eis dat de fit t.o.v. model 1 significant moet zijn verbeterd en ook dat het effect γ_{13} van een 0-effect te onderscheiden moet zijn, dan verkiezen model 2 boven model 1 in de volgende gevallen:

- grote versch:	blok 1;	resp. 2, 4 en 6	blok 7;	---
- kleine versch:	blok 1;	resp. 1, 4, 6 en 7	blok 7;	resp. 6
- alle versch:	blok 1;	resp. 1. 6 en 7	blok 7;	resp. 5

We zullen ons nu wenden tot de analyse met model 3. De resultaten hiervan zijn te vinden in tabel 4. In de eerste plaats valt op dat van de onderzochte modellen model 3 het minst vaak voldoet. Waar de complete reeks is geanaliseerd zelfs voor geen enkele respondent. Ook over respondenten gerekend waren de resultaten mager en voldeed het model voor geen enkele reeks, noch in blok 1, noch in blok 7. De fit van model 3 was vrijwel overal duidelijk slechter dan die van model 1. Alleen in blok 1, bij de kleine verschillen was de fit vaak beter dan die van model 1 en werd T over de respondenten gehalveerd. Gemiddeld gezien ging in vergelijking met model 1 de verklaarde variantie alleen vooruit in de reeks met de kleine stimulusverschillen in blok 1. Voor deze reeks in blok 7 stond die op hetzelfde niveau, maar in alle andere reeksen ging de r^2 achteruit. Vergeleken met model 2 gaven de resultaten een vergelijkbaar beeld. Alleen voor de reeks met de kleine verschillen stond de verklaarde variantie op gelijk niveau, in zowel blok 1 als blok 7, maar elders werd het niveau van model 2 nooit overtroffen. Uit het voorafgaande blijkt dat model 3 - zoals verwacht - slechts een alternatief kan zijn voor de andere modellen bij kleine stimulus verschillen.

Opvallend is dat model 3 in de reeks met de kleine verschillen in blok 7 betere resultaten te zien gaf dan in blok 1, terwijl juist in blok 7 het probleem van seriële effecten in de response in veel minder sterke mate leek te spelen.

Tabel 4: Resultaten model 3 df=2 (gestandariseerde oplossing)

	Blok 1					
resp.	γ_{11}	$-\gamma_{12}$	γ_{13}	r^2	T	p
alle versch. n=29						
1	1.00	1.04	1.11	.93	17.74	.00*
2	.92	.92	.98	.95	14.57	.00*
3	.88	.88	.95	.93	9.27	.01*
4	.93	.99	1.08	.89	11.31	.00*
5	.86	.86	.93	.86	14.66	.00*
6	.87	.94	1.05	.90	13.90	.00*
7	.79	.75	.90	.83	12.42	.00*
8	.86	.86	.93	.11	13.37	.00*
9	.98	.98	1.00	.97	10.24	.01*
10	.91	.86	.92	.95	16.68	.00*
T_{tot} df=20					134.26*	
grote versch. n=12						
1	1.02	1.09	1.19	.93	2.51	.29
2	.99	.99	.99	.97	7.35	.03*
3	.83	.83	.92	.92	5.83	.05
4	.93	.93	1.02	.86	15.89	.00*
5	.85	.85	.91	.85	10.01	.01*
6	.97	1.04	1.17	.87	7.21	.03*
7	.76	.68	.79	.79	1.50	.47
8	.81	.81	.85	.80	21.89	.00*
9	1.00	1.00	1.06	.98	3.26	.20
10	.93	.83	.88	.94	12.64	.00*
T_{tot} df=20					88.09*	
kleine versch. n=12						
1	.49	.49	.98	.96	10.98	.00*
2	.82	.82	1.01	.99	10.03	.01*
3	1.26	1.26	.95	.95	3.46	.18
4	.69	.69	1.01	.95	.84	.66
5	.79	.79	1.03	.99	.11	.95
6	1.28	1.29	1.02	.95	14.77	.00*
7	1.57	1.59	1.07	.83	15.61	.00*
8	1.07	1.07	1.02	.99	3.06	.22
9	.94	.94	.96	.98	7.25	.03*
10	.87	.88	1.00	.98	5.64	.06
T_{tot} df=20					71.74*	

(kritieke waarde bij $\alpha = 5\%$ χ^2_{20} : 31.41 χ^2_2 : 5.99 / * = T is significant)

Tabel 4, vervolg: Resultaten model 3 df=2 (gestandariseerde oplossing)

Blok 7

resp.	γ_{11}	$-\gamma_{12}$	γ_{13}	r^2	T	p
alle versch. n=29						
1	.98	.98	.99	.97	26.27	.00*
2	.99	.99	1.00	1.00	17.30	.00*
3	.94	.94	.97	.96	19.66	.00*
4	1.02	1.02	1.02	.98	18.62	.00*
5	.98	.98	1.01	.99	24.14	.00*
6	.96	.96	.97	.94	13.62	.00*
7	.97	.97	.98	.98	20.26	.00*
8	.90	.90	.95	.91	18.57	.00*
9	1.00	1.00	1.00	.99	11.39	.00*
10	.99	.99	.99	.99	19.50	.00*
T_{tot} df=20					189.33*	
grote versch. n=12						
1	.98	.98	.98	.95	16.55	.00*
2	.99	.99	.98	1.00	23.40	.00*
3	.96	.96	.99	.99	3.45	.18
4	1.04	1.04	1.05	.98	12.19	.00*
5	.99	.99	1.03	.99	43.92	.00*
6	.92	.92	.94	.93	10.17	.01*
7	.96	.96	.97	.98	8.17	.02*
8	.86	.86	.94	.85	19.90	.00*
9	.98	.98	.98	1.00	5.69	.06
10	.99	.99	.97	.99	15.50	.00*
T_{tot} df=20					158.93*	
kleine versch.n=12						
1	.73	.73	.99	.99	5.82	.05
2	.92	.92	1.00	1.00	1.87	.39
3	.97	.97	.98	.99	3.1	.21
4	.96	.96	.98	.98	8.25	.02*
5	1.02	1.02	.99	.99	5.99	.05
6	.74	.74	1.00	.98	1.05	.59
7	.98	.98	.99	.98	7.47	.02*
8	1.14	1.14	.96	.95	6.07	.05*
9	.90	.90	.99	.99	3.01	.22
10	.90	.90	1.00	1.00	2.48	.29
T_{tot} df=20					45.10*	

(kritieke waarde bij $\alpha = 5\%$ χ^2_{20} : 31.41 χ^2_2 : 5.99 / * = T is significant)

Een mogelijke verklaring zou kunnen zijn dat juist in blok 7, waar de respondenten door de standaard minder vrij waren in hun beantwoording, men probeert de beoordelingstaak te vereenvoudigen door de response op een voorafgaand stimulus als standaard te gebruiken. Met name daar waar de voorafgaande stimulus méér op de voorliggende stimulus lijkt dan de standaard. De voorafgaande stimulus verdringt zo de in de vraagstelling opgenomen standaard. Om een beter overzicht te krijgen over de modellen is een rangordening gemaakt van alle passende modellen, volgens de criteria die we reeds eerder hebben opgesteld, t.w.

- verkies model 2 boven model 1 als een significante verbetering in fit werd verkregen en ook γ_{13} significant is.
- verkies model 3 boven model 2 indien fit en verklaarde variantie tenminste gelijk zijn. De extra vrijheidsgraad geeft in geval van gelijke fit en dezelfde verklaarde variantie de doorslag.
- verkies model 1 boven model 3 als fit en verklaarde variantie tenminste gelijk zijn. In geval van vergelijkbaarheid geeft de grotere eenvoud van model 1 de doorslag.

De zo verkregen rangordening van de modellen is weergegeven in tabel 5. Voor blok 1 is de situatie het minst duidelijk. Voor de gehele reeks en de reeks met de grote stimulusverschillen lijkt model 1 het meest plausibel en biedt model 2 het alternatief. Voor

tabel 5: rangordening van de modellen 1,2 en 3 (1e interview)

	Blok1 1			Blok 7		
resp.	alle	grote	kleine	alle	grote	kleine
1	2	1,2,3	2	1	1	1,3
2	1	2,1	-	1	1	1,3
3	-	1,3	3	1	1,3	3
4	-	2	3,2	1	1	1
5	1	1	3	2	-	1,3
6	2	2,1	2	1	1	3,2
7	2	1,3	2	1	1	1
8	1	-	3	1	-	1
9	1	1,3	-	1	1,3	1,3
10	1	1	3	1	1	3

de reeks met de kleine stimulusverschillen voldoet model 1 in geen van de gevallen. Kennelijk is hier het risico van seriële afhankelijkheden in de responses het grootst, hetgeen in overeenstemming is met de bevindingen van Green, Luce en Duncan. Voor het modelleren van de seriële effecten bij de kleine stimulus verschillen lijkt model 3 het meest voor de hand te liggen, maar ook model 2 past in individuele gevallen. In de twee andere reeksen is model 3 echter nergens de eerste keus.

In blok 7, waar een standaard in de vraagstelling was aangeboden, is model 1 duidelijk het overwegende model. Het is in veel gevallen het enige- en in vrijwel alle overige gevallen het beste model. We hadden eerder reeds geconstateerd dat het gevaar voor seriële effecten in de response sterk afneemt als een standaard wordt gegeven. Door bovenstaande resultaten worden we gesterkt in deze mening. De resultaten van de complete reeks in blok 7 zouden de indruk kunnen wekken dat hier geen sprake meer zou zijn van seriële effecten. Dit is echter niet het geval, getuige het feit dat we wél effecten vinden in de subreeks met de kleine stimulus verschillen.

Stabiliteit door de tijd

Om een indruk te krijgen over de stabiliteit van de hier gevonden resultaten, werden dezelfde analyses tevens op de tweede dataset verricht. Uit deze analyses ontstond een met de hiervoor reeds gerapporteerde resultaten overeenkomstig beeld. Seriële effecten deden zich ook hier hoofdzakelijk voor in de reeksen met de kleine stimulusverschillen. Ook was de positieve invloed van de standaard op de onderdrukking van seriële effecten en het verhogen van de verklaarde variantie weer duidelijk waarneembaar. In tabel 6 is weergegeven welke modellen in de verschillende geanaliseerde reeksen voldeden en hun rangordening.

Voor de reeksen met de kleine verschillen lijkt model 3 het beste te voldoen, met name in het blok met de standaard is dat hier duidelijker waar te nemen dan in het eerste interview. Voor de overige reeksen gaat de keus vooral tussen de modellen 1 en 2.

Verder valt in tabel 5 zowel als in tabel 6 op dat de variatie in de gevonden response modellen sterk afneemt wanneer een standaard in de vraagstelling was geïntroduceert. Door de standaard formuleert een aanmerkelijk groter deel van de respondenten hun responses volgens eenzelfde model.

tabel 6: rangordening van de modellen 1,2 en 3 (2e interview)

	Blok1 1			Blok 7		
resp.	alle	grote	kleine	alle	grote	kleine
1	-	1	3,2	1	-	3
2	1	-	3	1	-	-
3	1	-	-	1	1	3
4	2,1	1,3	3,2	1	1	3
5	1	2,1	1	1	1	3
6	-	1	3,2	1	1	1,3
7	-	1	2,1,3	1	1,2	1
8	2	2,1,3	-	1	-	-
9	1	1	-	1	1	3,1
10	1	1	-	1	1	1,3

Opvallend bij de resultaten van het tweede interview is ook het groter aantal gevallen waarin geen van de drie onderzochte modellen voldeed. Dit deed de verdenking rijzen, dat het tweede interview minder zorgvuldig zou zijn ingevuld dan het eerste. Analyse van de betrouwbaarheid van de antwoorden wees echter uit dat dit niet het geval was, zodat we moeten aannemen dat we te maken hebben met toevallige storingen in de responses, dan wel dat de response hier in werkelijkheid anders moet worden gemodelleerd.

Nu we over de gegevens van beide interviews beschikken, kunnen we voor elke reeks en per respondent vaststellen waar de gevonden modellen in de twee interviews dezelfde zijn. Dit is weergegeven in tabel 7.

Hier blijkt dat voor de reeks met alle en de reeks met de grote verschillen alleen model 1 stabiel is over de twee interviews. Stabiele modellen waarin met seriële effecten rekening is gehouden vinden we alleen terug bij de reeksen met de kleine verschillen. Met name in blok 7 is model 3 sterk overwegend. Model 2 vinden we in het tweede interview zelden op dezelfde plaats terug, alleen in de reeks met kleine verschillen in blok 1.

In blok 7 treffen we vaker als in blok 1 hetzelfde model op dezelfde plaats aan. Het aantal respondenten voor wie de antwoorden in elk van de twee interviews volgens één model kan worden beschreven neemt toe, wanneer een standaard in de vraagstelling werd opgenomen Het opnemen van een standaard in de vraagstelling blijkt dus de stabiliteit van de gebruikte modellen te vergroten.

tabel 7: stabiele modellen in de twee interviews

	Blok1 1			Blok 7		
resp.	alle	grote	kleine	alle	grote	kleine
1	-	1	2	1	-	3
2	1	-	-	1	-	-
3	-	-	-	1	1	3
4	-	-	3,2	1	1	-
5	1	1	-	-	-	3
6	-	1	3	1	1	3
7	-	1	2	1	1	1
8	-	-	-	1	-	-
9	1	1	-	1	1	1,3
10	1	1	-	1	1	3

We stellen vast, dat het aanbieden van een standaard in de vraagstelling zowel de vergelijkbaarheid van de response functies over de respondenten, maar ook de stabiliteit door de tijd (over de interviews) vergroot. Dit lijkt niet onlogisch, omdat de vrijheid van antwoorden in blok 1 groter is dan in blok 7. Tenslotte blijkt uit de resultaten, dat het geven van één standaard niet voldoende is om seriële effecten bij kleine stimulus verschillen te voorkomen.

Conclusies

In dit onderzoek hebben we door het onder verschillende condities vergelijken van drie modellen voor het beschrijven van de relatie tussen stimuli en de daarop verkregen responses geprobeerd inzicht te krijgen in mogelijk optredende seriële effecten in de beoordelingen. We hebben geconstateerd dat de gevonden responsfuncties per respondent, met de gevolgde meetprocedure en met de aard van de stimulusverschillen kunnen variëren.

Seriële effecten in de response vonden we voornamelijk daar waar de verschillen tussen opvolgende stimuli klein waren, hetgeen de bevindingen van Green et al. bevestigde. Het bleek mogelijk deze effecten voor een deel te vermijden door een standaard in de vraagstelling op te nemen. Voor zover de verschillen tussen opvolgende stimuli niet te klein zijn lijken de resultaten erop te wijzen dat seriële effecten door de standaard geheel verdwijnen en dat de relatie tussen stimuli en responses in die gevallen ook over de cases beschouwd met het eenvoudige model 1 kan worden beschreven.

Bij kleine verschillen tussen opeenvolgende stimuli is het geven van een standaard echter niet voldoende voor het voorkomen van seriële effecten. Wanneer deze effecten in de responses optreden ondanks de gegeven standaard, lijkt men om tot een antwoord te komen uit te gaan van de response op de voorafgaande stimulus. Vervolgens corrigeert men voor het verschil tussen de twee stimuli. De standaardstimulus wordt zo als het ware vervangen door een stimulus die meer lijkt op het huidig voorliggend stimulus. Deze gang van zaken wordt het best weergegeven door model 3.

Door het geven van een standaard werd de vergelijkbaarheid over de respondenten van de bebruikte modellen, alsmede de stabieliteit van de manier van antwoorden door de tijd vergroot.

Om al te duidelijke risico's voor wat betreft seriële effecten in de responses te voorkomen en vergelijkbaarheid van antwoorden over respondenten te maximaliseren valt het aan te bevelen om kleine verschillen tussen opvolgende stimuli te vermijden. Daarnaast is het aanbieden van tenminste één standaard in de vraagstelling noodzakelijk.

Boomsma, A. (1983) *On the robustness of the Lisrel procedure (maximum likelihood estimation) against small and moderate sample size and non-normality.* Proefschrift R.U. Groningen / Amsterdam: Stichting voor Sociometrisch Onderzoek

Cross, D.V. (1973) Sequential dependencies and regression in psychophysical studies. *Perception & Psychophysic, 14*, 547-552

Doorn, L.J. van, Saris, W.E. & Lodge, M. (1983). Discrete or continuous measurement. What difference does it make ? / *Kwantitatieve Methoden, 4*, (10), 104 - 121

Green, D.,Luce, R. & Duncan, J. (1977). Variability and sequential effects in magnitude production and estimation of auditory intensity. *Perception & Psychophysics, 22*, 450-45

Johnston, J. (1972) *Econometric Methods*, 2e ed. New York : MacGraw - Hill Book Company

Jöreskog, K.G. & Sörbom, D. (1978) *Lisrel IV. A general computer program for estimation of linear structural equation system by maximum likelihood methods.* / Chicago: International Educational Services

Lodge, M. (1981) *Magnitude scaling, quantative measurement of opinion.* Beverly Hills: Sage

Pijper, M. de. (1982) Percase. A computer program for individual analysis. Vakgroep Methoden en Technieken, Vrije Universiteit, Amsterdam

Pijper, M. de, Saris, W.E. & Gallhofer, I.N. (1986) User manual of the SRF interview program. / Amsterdam: Sociometric Research Foundation.

Saris, W.E. (1982) Different questions, different variables, in C. Fornell (ed), *A second generation of multivariate analysis.* / New York: Praeger Publishers.

Saris, W.E. et al. (1987) *Variation in response behaviour a source of measurement error.* Amsterdam: Sociometric Research Foundation.

Saris, W.E., Satorra, A. & Sörbom, D. (1987) Detection and correction of specification errors, in C. C. Clogg (ed) *Sociological Methodology* / San Francisco: Jossey - Bass

Satorra, A & Saris, W.E. (1982) The accuracy of a procedure for calculating the power of the like;ihood ratio test as used within the LISREL framework. In *Sociometric Research* / Amsterdam: Sociometric Research Foundation.

Stevens, S.S. (1975) *Psychophysics: Introduction to its perceptual, neural and social effects.* / New York: Wiley

Theil, H. (1971) *Principles of econometrics* / New York: Wiley

Wegener, B. (1980) Fitting category to magnitude scales for a dozen survey assessed attitudes. In B.Wegener (ed), *Social attitudes and psychophysical Measurement.* Hillsdale, N. J. : Erlbaum

Ontvangen: 10-06-86
Geaccepteerd: 06-10-86