De adequaatheid van covariantiestructuurmodellen: een overzicht van maten en indexen[1]

Anne Boomsma[2]

Samenvatting

Voor de analyse van covariantiestructuren wordt een overzicht gegeven van verschillende maten en indexen die gebruikt kunnen worden voor modelselectie en voor het evalueren van de passing van een model. Alle maten die het programma LISREL 8 berekent worden behandeld, naast enkele andere die van belang zijn voor een enigszins omvattend overzicht. Er wordt een poging gedaan de verschillende maten op een gestructureerde manier naast elkaar te zetten, waarbij de nadruk ligt op de statistische vraag die de onderzoeker met het gebruik van de verschillende typen indexen tracht te beantwoorden.

Inleiding

Er zijn in de literatuur de laatste jaren tal van passingsmaten en indexen gepubliceerd, die moeten helpen bij het het evalueren van een geschat model en bij het kiezen van een adequaat model uit een reeks geschatte modellen (het probleem van modelselectie). Tot voor kort was menig LISREL-gebruiker zich daar niet erg van bewust, omdat al die maten niet door het programma werden afgedrukt. Met LISREL 8 (Jöreskog & Sörbom, 1993a, 1993b) wordt er echter zoveel informatie over de passing van modellen gepresenteerd, dat het begrijpelijk wordt dat gebruikers door de bomen het bos niet meer zien; dat geldt overigens ook voor andere programma's op het gebied van de analyse van covariantiestructuren, zoals EQS (Bentler, 1989) en AMOS (Arbuckle, 1993). Het aanbod in programmatuur van het grote aantal indexen ter evaluatie van het model lijkt niet te worden geremd door enige mate van spaarzaamheid of

---

[1] Lezing voor de Sociaal-Wetenschappelijke Sectie van de Vereniging voor Statistiek, gehouden te Utrecht op 21 maart 1995.

[2] A. Boomsma, Vakgroep Statistiek & Meettheorie, Rijksuniversiteit Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen.

selectiviteit. 'Unfortunately there seems to be a proliferation of model selection tools, so that it seems likely we will need a tool to select model selection tools in the not too distant future.' (De Leeuw, 1990, p. 240) De klemmende vraag is daarom of in het geheel van dat aanbod enige structuur is aan te brengen: zijn de diverse maten te groeperen naar de vragen die ze geacht worden te beantwoorden, of naar de aspecten waarop de nadruk wordt gelegd? Alvorens op die vragen in te gaan schetsen we het probleem van modelselectie.

Het probleem van modelselectie

Een vaak voorkomende situatie is dat een onderzoeker niet de adequaatheid van één enkel model M_j wil inspecteren, maar dat hij een aantal rivaliserende of alternatieve modellen in ogenschouw neemt, zeg M_1, M_2, ..., M_m, die op theoretische gronden elk meer of minder acceptabel zijn. Voor een van de modellen M_j, $j=1,\ldots,m$, zou de onderzoeker graag min of meer definitief willen kiezen. We gaan er voor het gemak even vanuit dat er zich bij het schatten van elk model M_j geen onregelmatigheden voordoen, die direct tot verwerping van dat model zouden kunnen leiden, bij voorbeeld doordat er ontoelaatbare schattingen van modelparameters θ optreden.

Bij modelselectie spelen naast de globale passing van het model overwegingen van spaarzaamheid een rol (cf. Mulaik, James, Van Alstine, Bennett, Lind & Stillwell, 1989; Bentler & Mooijaart, 1989; McDonald & Marsh, 1990). Enerzijds moet de passing van het model niet al te slecht zijn, anderzijds moet vermeden worden dat het model door een toename van het aantal te schatten parameters te complex wordt, of triviaal wanneer het om een perfect passend model gaat. Om die redenen zijn er bij voorbeeld criteria ontwikkeld die gebruik maken van correctietermen voor het aantal modelparameters t_j, wel of niet in relatie tot de steekproefgrootte N, zoals het informatiecriterium van Akaike (1973, 1974, 1987) en dat van Schwarz (1978). Zulke correctietermen worden wel penalty-functies genoemd. Het algemene idee daarbij is dat er (relatieve) strafpunten worden uitgedeeld voor het toevoegen van te schatten parameters. Vanuit dat gezichtspunt beogen adequate selectiecriteria een passend compromis te vinden tussen een keuze voor een model met weinig parameters (spaarzaamheid) met gevaar voor onderpassing ('underfit') en een

model met veel parameters (complexiteit) met gevaar voor overpassing ('overfit'). Voor alle praktische doeleinden geeft de onderzoeker de voorkeur aan een comfortabel model, dat niet te strak en niet te ruim zit.

In de praktijk van modelselectie ligt de nadruk vaak op het ontwikkelen van een model op basis van de enig beschikbare steekproef, en wordt de selectie bovendien vaak ook niet gehinderd door stapsgewijze, steekproef-gestuurde modelmodificaties zonder enige vorm van kruisvalidatie achteraf. De houdbaarheid van een model staat en valt echter met zijn predictieve validiteit in nieuwe, onafhankelijke steekproeven uit dezelfde populatie. Cudeck & Browne (1983) leggen de klemtoon bij modelselectie fundamenteel op dit aspect, hetgeen leidt tot hun methodologische nadruk op vormen van kruisvalidatie. 'Selection methods should de-emphasize the practice of developing a model in a single sample of data, and instead should be concerned with identifying models which will perform optimally in future samples. One reasonable method for this purpose is cross-validation.' (o.c., pp. 150-151)

Verderop zullen we overigens zien dat er nauwe overeenkomsten bestaan tussen het gebruik van het informatiecriterium van Akaike en benaderingen die de nadruk op kruisvalidatie leggen.

Bij informatie-, kruisvalidatie- en andere criteria berekent men deze maten bij elk van de modellen M_j, $j=1,\ldots,m$. Vervolgens brengt men een rangorde in de modellen aan middels het gekozen selectiecriterium, waarna men voor het model kiest met de kleinste of grootste criteriumwaarde.

De informatiematen en andere modelindexen voor een model M_j worden ook vaak vergeleken met twee andere modellen, die als twee uitersten in een keten van wel of niet geneste modellen kunnen worden beschouwd. We noemen ook nog een derde 'model', dat soms een rol bij passingsindexen speelt.

a. Het **onafhankelijkheidsmodel M_i** ('independence model') is een model waarbij de k geobserveerde variabelen niet met elkaar zijn gecorreleerd. Als deze variabelen een multivariate normale verdeling hebben betekent dit dat de variabelen statistische onafhankelijk zijn; dit impliceert dat de populatiecovariantiematrix $\Sigma(\theta)$ een (k x k) matrix is met nullen buiten de diagonaal. Bij zo'n model voor onafhankelijke of ongecorreleerde variabelen, aangeduid als M_i, moeten er $t_i = k$ parameters worden geschat, namelijk de varianties van de k geobserveerde variabelen. Het aantal

vrijheidsgraden van M_i is gelijk aan $d_i = k(k+1)/2 - k = k(k-1)/2$. In model M_i bestaan er geen causale verbanden tussen de variabelen. De onderzoeker pretendeert daarmee als het ware dat hij niets weet over de gerichte samenhangen tussen de variabelen.

Het onafhankelijkheidsmodel fungeert in de literatuur vaak als een basismodel ('baseline model') M_b, of nulmodel ('null model') M_0, waartegen andere modellen worden afgezet. Het is het eenvoudigste, meest restrictieve model dat als standaard voor modelvergelijkingen met minder restrictieve modellen M_j kan worden gebruikt. Sobel & Bohrnstedt (1985) hebben het gebruik van het onafhankelijkheidsmodel als een niet-informatief basismodel aan kritiek onderworpen; zie ook Mulaik et al. (1989).

b. Het **verzadigde model M_s** ('saturated model') is een model dat perfect bij de steekproefgegevens past. Het aantal te schatten parameters van M_s is $t_s = k(k+1)/2$, zodat het aantal vrijheidsgraden d_s gelijk is aan nul. Dit is een model waarbij de causale verbanden een perfecte, maar niet noodzakelijkerwijs unieke, representatie van de steekproefgegevens opleveren. De onderzoeker pretendeert dat hij (vrijwel) alles over de gerichte samenhangen tussen variabelen weet. Het is een triviaal model, omdat vrijwel ieder model perfect passend is te krijgen door voldoende parameters vrij te laten.

c. Het **ultieme nulmodel M_u** relateren we aan de situatie waarbij er geen parameterschattingen worden gemaakt, om preciezer te zijn aan die waarbij alle schattingen gelijk aan nul worden gesteld. Een gepostuleerd model M_j kan worden vergeleken met dit nulmodel M_u. Het verschil tussen beide geeft dan de winst aan van het schatten van een model ten opzichte van het in het geheel niet schatten daarvan, in de strikte betekenis van het fixeren van parameters op een vaste waarde van nul. Deze vergelijking speelt bij voorbeeld bij de 'goodness-of-fit'-index (GFI) een rol. Het aantal vrijheidsgraden van het nulmodel M_u is gelijk aan $d_u = k(k+1)/2$, omdat $t_u = 0$.

Is het gepostuleerde model geldig? Een toets voor exacte passing

Om de passing van een model te beoordelen zullen veel onderzoekers allereerst kijken naar de waarde van de χ^2-toetsingsgrootheid, die we als een

globale passingsmaat kunnen beschouwen. Om de vraag te beantwoorden of een model geldig of waar is, zo men wil, ('whether the model holds'), hebben we een formele notatie nodig (cf. Browne & Cudeck, 1992; Jöreskog & Sörbom, 1993a, p. 117), die we eerst beknopt, later iets genuanceerder introduceren.

We nemen aan dat de steekproefcovariantiematrix **S** met toenemende steekproefgrootte N in waarschijnlijkheid convergeert naar de populatiecovariantiematrix $\boldsymbol{\Sigma}_0$. Als $\boldsymbol{\theta}_0$ nu de waarde van $\boldsymbol{\theta}$ is die de algemene **doelfunctie** $F[\boldsymbol{\Sigma}_0;\boldsymbol{\Sigma}(\boldsymbol{\theta})]$ minimaliseert, dan zeggen we dat het model waar is als geldt dat $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. Maar in de praktijk hebben we uitsluitend te maken met een steekproefcovariantiematrix **S**, die gebaseerd is op een eindige steekproefomvang N. In dat geval is $\hat{\boldsymbol{\theta}}$ de waarde van $\boldsymbol{\theta}$ die de doelfunctie $F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})]$ minimaliseert. Om te toetsen of een model geldig is, gebruiken we, onder de aanname dat het model in de populatie ook werkelijk geldig is, een asymptotisch chi-kwadraat-verdeelde toetsingsgrootheid, die voor de algemene doelfunctie $F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})]$ is gedefinieerd als

$$\chi^2 = n\hat{F} = (N-1)F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})] \quad . \tag{1a}$$

Als we bij modelselectiecriteria expliciet willen aangeven dat het om model M_j gaat kunnen we (1a) ook schrijven als

$$\chi^2_j = n\hat{F}_j = (N-1)F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)] \quad . \tag{1b}$$

Let wel dat $\chi^2_j = n\hat{F}_j$ hier betekent de χ^2-waarde die bij de schatting van model M_j hoort, niet een χ^2-waarde onder een chi-kwadraat-verdeling met j vrijheidsgraden.

De minimum waarde van de algemene doelfunctie in (1a) en (1b),

$$\hat{F} = \hat{F}_j = F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)] \quad , \tag{2}$$

wordt wel de minimum waarde van de **steekproefdiscrepantiefunctie** genoemd. Ze wordt door LISREL 8 afgedrukt onder de term 'minimum fit function value'.

Onder de assumptie van multivariaat-normaal-verdeelde geobserveerde variabelen, geldt voor de meest aannemelijke schattingmethode met doelfunctie $F_{ML}[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_0)]$, onder de additionele aanname dat het gepostuleerde model waar is, dat asymptotisch, in grote steekproeven, de statistische grootheid $(N-1)F_{ML}[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_0)]$ een centrale χ^2-verdeling heeft met $d = k(k+1)/2 - t$ vrijheidsgraden, waarbij k het aantal geobserveerde variabelen is en t het aantal onafhankelijke parameters dat geschat moet worden. Onder de normaliteitsassumptie geldt dat ook voor de doelfunctie van de GLS-schattingsmethode ('generalized least squares'), en onder de aanname dat de correcte wegingsmatrix $\mathbf{W}$ wordt gebruikt eveneens voor de WLS-schattingsmethode ('weighted least squares'). De algemene doelfunctie $F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})]$ in (1a) slaat daarom in het vervolg op deze drie gevallen.

Voor de basismodellen M_i en M_s kan de χ^2-waarde vergelijkenderwijs ook worden uitgerekend. De χ^2-waarde die is geassocieerd met het model voor onafhankelijke of ongecorreleerde variabelen, kortweg model M_i, duiden we voor de algemene doelfunctie F aan als

$$\chi_i^2 = n\hat{F}_i = (N-1)F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_i)] \quad . \tag{3}$$

Voor het verzadigde model M_s met een perfecte passing geldt uiteraard

$$\chi_s^2 = n\hat{F}_s = 0 \quad . \tag{4}$$

De overschrijdingskans dat de stochastische variabele $n\hat{\underline{F}}_j$ een waarde aanneemt die groter is dan de gevonden waarde $n\hat{F}_j$ geven we aan als

$$P\{n\hat{\underline{F}}_j > n\hat{F}_j\} \quad . \tag{5}$$

Dit is de zogenaamde P-waarde van $n\hat{F}_j$ onder een centrale χ^2-verdeling met d_j vrijheidsgraden, die door LISREL 8 wordt afgedrukt.

Voor een verantwoord gebruik van de χ^2-toetsingsgrootheid moet er steevast van uit worden gegaan dat het gepostuleerde model M_j in de populatie geldig is. De statistische aanname is dat het model exact geldig is in de populatie. Vandaar dat Browne & Cudeck (1992, p. 240) in dit verband spreken van een toets voor **exacte passing**. Maar hoe realistisch is deze aanname van een exact geldend model eigenlijk?

De rol en de plaats van het ware populatiemodel

Voor een aantal maten die we verderop tegenkomen is het nuttig kennis te nemen van Cudeck & Henly (1991) en van Browne & Cudeck (1992). De positie van deze auteurs ten aanzien van het modelleren van covariantiestructuren wordt geïllustreerd door de volgende opinies en stellingnames.

'Yet no model is completely faithful to the behavior under study. Models usually are formalizations of processes that are extremely complex. It is a mistake to ignore either their limitations or their artificiality. The best one can hope for is that some aspect of a model may be useful for description, prediction, or synthesis. The extent to which this is ultimately successful, more often than one might wish, is a matter of judgment.' (Cudeck & Henly, 1991, p. 512)

'In applications of the analysis of covariance structures in the social sciences it is implausible that any model that we use is anything more than an approximation to reality. Since a null hypothesis that a model fits exactly in some population is known a priori to be false, it seems pointless to even try to test whether it is true.' (Browne & Cudeck, 1992, p. 231)

En zelfs als we zouden aannemen dat een model in de populatie zou gelden, dan nog is het in het algemeen onredelijk te veronderstellen dat de functie $\Sigma(\boldsymbol{\theta}_0)$ in de praktijk bekend zal zijn. Integendeel, zo stellen Cudeck & Henly (1991, p. 513): 'In fact, most often the purpose of an investigation is simply to summarize the data, and not to discover what the operating model might be. It still is useful to suppose, as do all researchers who study quantitative models, that one or more simplified or idealized structures provide some kind of representation of the unknown process that generates

Σ_0.' Zij noemen deze structuren $\Sigma_j = \Sigma(\theta_j)$, j=1,...,m, dan ook '**approximating models**'.

Zij sluiten daarbij aan bij opvattingen die eerder bij herhaling door De Leeuw (1983, 1988, 1990) naar voren zijn gebracht. 'Of course it is generally true that models are approximations. ... Models are never true, everything is significant if the sample size is large enough. That is as it should be, and it should not prevent us from trying to make our models as realistic as possible.' (De Leeuw, 1983, p. 115) 'Models are never true, they are at best good approximations, where 'good' means good enough for practical purposes.' (De Leeuw, 1988, p. 120)

De vraag of een model in de populatie geldig is wordt door de geciteerde auteurs feitelijk als onzinnig terzijde gezet. De vraag naar de correctheid van het model lijkt dus niet aan de orde. Bovendien geldt dat wanneer de steekproefomvang voldoende groot is, zelfs modellen die de situatie in de populatie redelijk dicht benaderen verworpen zullen worden. Het lijkt daarom veel beter de mate van gebrek aan passing ('lack-of-fit') van een model te schatten. In welke mate schiet het model tekort? Oftewel, hoe groot is de **deficiëntie** van een model? Zo zouden we ook de deficiëntie kunnen vergelijken van een aantal mogelijke modellen M_j, die elk als **benaderingen** van de werkelijkheid worden opgevat.

Het zou interessant zijn na te gaan hoe de hier geschetste ideeën aansluiten bij wetenschapstheoretische opvattingen over waarheidsbenadering ('truthlikeness'); zie bij voorbeeld Kuipers (1992).

Soorten fouten en discrepanties

Ter formalisering van hun opvattingen over de relaties tussen de toestand of het proces in de populatie, de onmogelijkheid die toestand of dat proces exact te bepalen, en het postuleren van benaderingsmodellen met een beperkte doelstelling, bouwen Cudeck & Henly (1991) voort op het werk van Linhart & Zucchini (1986). Ze onderscheiden daarbij drie soorten matrixen.

1. Σ_0 is de populatiecovariantiematrix, die, zoals hierboven aangeduid, in het algemeen niet overeenstemt met het model dat door onderzoekers wordt geschat of gepostuleerd, omdat ze altijd een benaderingsmodel gebruiken. De matrix $\Sigma_0 = \Sigma(\theta_0)$ wordt ook wel het operationele model ('operating

model') genoemd (zie Cudeck & Henly, 1991, p. 513).

2. $\tilde{\Sigma}_j = \Sigma(\tilde{\theta}_j)$ representeert de beste passing van het gepostuleerde model M_j bij de populatiecovariantiematrix Σ_0 in termen van de geselecteerde doelfunctie F. Dat wil zeggen $\tilde{\theta}_j = \arg\min F[\Sigma_0;\Sigma(\theta_j)]$.
3. $\hat{\Sigma}_j = \Sigma(\hat{\theta}_j)$ representeert de beste passing van model M_j bij de steekproefcovariantiematrix $\mathbf{S}$. Dat wil zeggen $\hat{\theta}_j = \arg\min F[\mathbf{S};\Sigma(\theta_j)]$.

Terwijl zowel Σ_0 als $\tilde{\Sigma}_j$ onbekende, gefixeerde matrixen zijn, is $\hat{\Sigma}_j$ een stochastische matrix, die van steekproef tot steekproef varieert.

De matrix $\Sigma(\tilde{\theta}_j)$ is een benadering van Σ_0 in de populatie, terwijl de matrix $\Sigma(\hat{\theta}_j)$ een schatting is van Σ_0 bij een gegeven steekproefcovariantiematrix $\mathbf{S}$ (Cudeck & Henly, 1991, p. 514). Als de steekproefomvang $N \to \infty$ dan $\mathbf{S} \to \Sigma_0$, dat wil zeggen de steekproefcovariantiematrix convergeert in waarschijnlijkheid naar Σ_0. Tegelijkertijd geldt, als $N \to \infty$ dan $\hat{\Sigma}_j \to \tilde{\Sigma}_j$, dat wil zeggen de matrix $\hat{\Sigma}_j$ convergeert in waarschijnlijkheid naar $\tilde{\Sigma}_j$.

Bij het kijken naar verschillen tussen de drie genoemde matrixen onderscheiden Browne & Cudeck (1992) nu drie soorten fouten, die middels drie soorten discrepanties F (Cudeck & Henly, 1991) worden geschat.

a. **De benaderingsfout**

De benaderingsfout ('error of approximation') verwijst naar het gebrek aan passing van een model M_j bij de populatiecovariantiematrix Σ_0, waarbij gekeken wordt naar verschillen tussen Σ_0 en $\tilde{\Sigma}_j$. Ze speelt een rol bij de vraag hoe goed een model M_j met onbekende, maar optimaal gekozen parameterwaarden bij de eveneens onbekende populatiecovariantiematrix Σ_0 past. Een maat voor de benaderingsfout is

$$F_0 = F(\Sigma_0;\tilde{\Sigma}_j) \quad , \tag{6}$$

de **benaderingsdiscrepantie** ('discrepancy due to approximation'), waarbij F een gekozen doelfunctie is. Het is een maat voor de ongelijkheid tussen Σ_0 en een benaderingsmodel M_j in de populatie. Daarom noemen Jöreskog & Sörbom (1993a, p. 123) de grootheid F_0 de waarde van de 'population discrepancy function' (PDF). De benaderingsdiscrepantie is een onbekende constante, die in het algemeen kleiner wordt als er parameters aan het

model worden toegevoegd. De benaderingsfout is een niet-stochastische fout, die niet van de steekproefomvang afhangt.

b. **De schattingsfout**

De schattingsfout ('error of estimation') verwijst naar verschillen tussen $\bar{\Sigma}_j$ en $\hat{\Sigma}_j$. Een maat voor de schattingsfout is

$$F_e = F(\bar{\Sigma}_j ; \hat{\Sigma}_j) \quad , \tag{7}$$

de **schattingsdiscrepantie** ('discrepancy due to estimation'), welke een stochastische variabele is. De verwachte waarde van F_e is bij benadering gelijk aan t/n, waarbij t het aantal te schatten parameters is en n = N-1 (zie Browne & Cudeck, 1992, p. 236); dus

$$E\{F_e\} \approx t/n \quad . \tag{8}$$

De verwachte waarde van F_e neemt toe als het aantal parameters groter wordt en de steekproefomvang gelijk blijft.

c. **De totale fout**

De totale fout ('overall error') verwijst naar verschillen tussen Σ_0 en $\hat{\Sigma}_j$. Ze speelt een rol bij de vraag hoe goed een model M_j met parameterschattingen op basis van een steekproefcovariantiematrix **S** bij de onbekende populatiecovariantiematrix Σ_0 past. Een maat voor de totale fout is

$$F_t = F(\Sigma_0 ; \hat{\Sigma}_j) \quad , \tag{9}$$

de **totale discrepantie** ('overall discrepancy'), welke een stochastische variabele is. De verwachte waarde van F_t is bij benadering gelijk aan de som van de benaderingsfout F_0 en de verwachte waarde van de schattingsfout F_e (zie Shapiro, 1985; Browne & Shapiro, 1988):

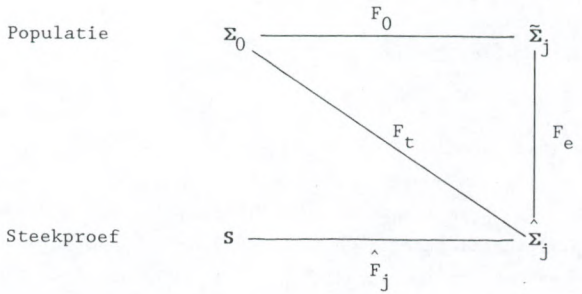
$$E\{F_t\} \approx F(\Sigma_0;\tilde{\Sigma}_j) + E\{F_e\} \approx F_0 + t/n \quad . \tag{10}$$

De totale fout is een stochastische fout, die wel van de steekproefomvang afhangt.

Ter onderscheiding van de drie discrepanties (6), (7) en (9) zij opgemerkt dat eerder [zie uitdrukking (2)] bij een gegeven steekproefcovariantiematrix **S**, de stochastische grootheid $\hat{F}_j = F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)]$ de **steekproefdiscrepantie** werd genoemd.

Als de steekproefomvang $N \to \infty$, dan gaan de drie stochastische variabelen $F_e \to 0$, $F_t \to F_0$ en $\hat{F}_j \to F_0$. Vergelijk Cudeck & Henly (1991, p. 514), die de onderlinge afstanden tussen de te onderscheiden matrixen ongeveer weergeven zoals in Figuur 1.

Zonder daarop nader in te gaan, is het interessant te wijzen op overeenkomsten en accentverschillen tussen de ideeën die aan Figuur 1 ten grondslag liggen en die welke door De Leeuw (1988, p. 124f.) naar voren zijn gebracht.



Figuur 1. Discrepanties tussen de populatiecovariantiematrix Σ_0, de steekproefcovariantiematrix **S**, de populatiebenadering $\tilde{\Sigma}_j$ van Σ_0, en de steekproefschatting $\hat{\Sigma}_j$ van Σ_0 voor model M_j. De afstanden tussen deze matrixen worden weergegeven door de benaderingsdiscrepantie F_0, de schattingsdiscrepantie F_e, de totale discrepantie F_t, en de steekproefdiscrepantie $\hat{F}_j$.

In het modelleringsconcept, zoals hierboven weergegeven, wordt ervan uitgegaan dat een willekeurig model M_j niet precies geldig is. Iedere onderzoeker zit er bij zijn modelkeuze enigszins naast, enigszins bezijden de waarheid. Het zou daarom mooi zijn als we een indicatie zouden kunnen krijgen van de mate waarin de onderzoeker naast de waarheid zit. Daarvoor kunnen we naar twee statistische grootheden kijken: naar de zogenaamde niet-centraliteitsmaat en naar de benaderingsfout.

De niet-centraliteitsparameter en de benaderingsfout

Om te verdisconteren dat model M_j niet exact geldig is in de populatie, kijken we eerst naar een schatting van de zogenaamde niet-centraliteitsparameter λ en naar betrouwbaarheidsintervallen voor λ. Zoals zal blijken hangt de benaderingsfout [zie uitdrukking (6)] daar zeer nauw mee samen.

Als het gepostuleerde model niet geldig is, als dus geldt dat $\Sigma_0 \neq \Sigma(\boldsymbol{\theta}_0)$ of $F_0 > 0$, dan heeft de statistische grootheid $n\hat{F}_j$ [zie uitdrukking (1b)] asymptotisch een **niet-centrale χ^2-verdeling** met d_j vrijheidsgraden en **niet-centraliteitsparameter** λ, die (cf. Browne, 1984, p. 69) gedefinieerd is als

$$\lambda = (N-1)F[\Sigma_0;\tilde{\Sigma}(\boldsymbol{\theta}_j)] = nF_0 \quad . \tag{11}$$

De niet-centraliteitsparameter is dus een functie van de steekproefomvang N en de benaderingsfout F_0. Om een schatting te krijgen van de benaderingsfout kunnen we niet de minimum waarde $\hat{F}_j = F[\mathbf{S};\Sigma(\hat{\boldsymbol{\theta}}_j)]$, de minimum waarde van de **steekproefdiscrepantiefunctie** gebruiken, omdat $\hat{F}_j$ geen zuivere schatter is van de benaderingsdiscrepantie $F_0 = F(\Sigma_0;\tilde{\Sigma}_j)$; zie Browne & Cudeck (1992, p. 237). Maar hoe groot is de verwachte waarde van $\hat{F} = \hat{F}_j$ eigenlijk?

De verwachtingswaarde van een stochastische grootheid met een niet-centrale χ^2-verdeling met d vrijheidsgraden en niet-centraliteitsparameter λ, is gelijk aan $\lambda + d$. Met (1a) en (11) volgt daaruit dat de verwachte waarde van $\hat{F}$ benaderd kan worden als

$$E\{\hat{F}\} \approx F_0 + d/n = F_0 + (k^* - t)/n = F_0 + k^*/n - t/n \quad , \tag{12}$$

waarbij $n = N-1$, $k^* = k(k+1)/2$ en t het aantal te schatten modelparameters is (cf. McDonald, 1989, p. 99). Dit impliceert dat de grootheid

$$\hat{\lambda}^* = \hat{F} - d/n = (n\hat{F} - d)/n \tag{13}$$

een zuiverder schatter voor de benaderingsfout F_0 is dan $\hat{F}$. We gebruiken in (13) de notatie $\hat{\lambda}^*$, omdat deze grootheid ook een schatter van de herschaalde niet-centraliteitsparameter $\lambda/n = F_0$ is (cf. McDonald, 1989, p. 99, formule 5).

Door steekproeffluctuaties zou (13) een negatieve waarde kunnen aannemen. Daarom gebruiken Browne & Cudeck (1992, p. 237) de volgende puntschatter voor de benaderingsdiscrepantie:

$$\hat{F}_0 = \max\{\hat{F} - d/n, 0\} = \max\{F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})] - d/n, 0\} \quad . \tag{14}$$

Uit (11) en (14) volgt dat een puntschatting van de niet-centraliteitsparameter λ wordt verkregen met de schatter

$$\hat{\lambda} = \max\{n\hat{F} - d, 0\} = \max\{nF[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})] - d, 0\} \quad . \tag{15}$$

Vergelijk McDonald (1989, p. 99, formule 6), die ook nog een transformatie van (13), de schatter van de herschaalde niet-centraliteitsparameter $\lambda^* = \lambda/n$, introduceerde (o.c., formule 8):

$$MCI = \exp\{-\hat{\lambda}^*/2\} = \exp\{-[(n\hat{F} - d)/n]/2\} \quad . \tag{16}$$

Deze index wordt door Hu & Bentler (1995, p. 86) McDonald's centraliteitsindex genoemd. De waarde van MCI ligt doorgaans tussen nul en één, maar ze kan door steekproeffluctuaties een waarde groter dan één aannemen. Om dat laatste te vermijden stellen we voor op basis van (14) de grootheid

$$MNCI = \exp\{-\hat{F}_0/2\} = \exp\{-[\max(\hat{F} - d/n, 0)]/2\} \qquad (17)$$

te gebruiken, die we McDonald's genormeerde centraliteitsindex noemen.

Hoe kleiner $\hat{F}_0$ en $\hat{\lambda}$ des te kleiner is de benaderingsfout, en des te groter kan het vertrouwen zijn in de validiteit van het gebruik van de centrale χ^2-toetsingsgrootheid voor exacte passing.

Om een indruk te krijgen van de variabiliteit van de schattingen van F_0 en λ kunnen we op basis van vroeg werk van Steiger & Lind (1980) betrouwbaarheidsintervallen voor deze populatiegrootheden schatten.

We bespreken eerst hoe zulke intervallen voor λ kunnen worden geschat. Laat $G(x|\lambda,d)$ de verdelingsfunctie van de niet-centrale χ^2-verdeling zijn met parameters λ en d. Bij gegeven $x = n\hat{F}$, d en een betrouwbaarheidscoëfficiënt van $1 - \alpha$ kunnen we een $100(1-\alpha)$%-betrouwbaarheidsinterval voor λ uitrekenen door de volgende twee vergelijkingen op te lossen, respectievelijk voor de benedengrens λ_L en de bovengrens λ_U van het interval:

$$G(n\hat{F}|\lambda_L,d) = 1 - \alpha/2 \quad , \qquad (18a)$$

$$G(n\hat{F}|\lambda_U,d) = \alpha/2 \quad . \qquad (18b)$$

De gevonden waarden uit deze twee vergelijkingen,

$$(\hat{\lambda}_L, \hat{\lambda}_U) \quad , \qquad (19)$$

zijn schattingen van de grenzen van het $100(1-\alpha)$%-betrouwbaarheidsinterval voor λ. Zie Browne & Cudeck (1992, p. 238) voor details.

Uit de gelijkheid $F_0 = \lambda/n$ [zie (11)] volgt dat een schatting van het $100(1-\alpha)$%-betrouwbaarheidsinterval voor F_0, bij berekende schattingen $\hat{\lambda}_L$ en $\hat{\lambda}_U$, wordt gegeven door

$$(\hat{\lambda}_L/n,\ \hat{\lambda}_U/n) \quad . \tag{20}$$

De RMSEA als functie van de benaderingsfout

Omdat de waarde van F_0, de benaderingsdiscrepantie, in het algemeen afneemt als er parameters aan een model worden toegevoegd kan het de voorkeur verdienen voor het aantal vrijheidsgraden d te corrigeren. Dan wordt gebruik gemaakt van de zogenaamde 'Root Mean Square Error of Approximation' (RMSEA), een maat voor de discrepantie per vrijheidsgraad, die door Steiger (1990) is aanbevolen; zie ook Steiger & Lind (1980). De RMSEA (oftewel de wortel uit de gemiddelde gekwadrateerde benaderingsfout) is gedefinieerd als

$$\text{RMSEA} = (F_0/d)^{1/2} \quad . \tag{21}$$

Merk op dat de RMSEA niet stochastisch is, ze geeft een indicatie van de benaderingsfout, welke geen stochastische fout is.

Wat betreft de interpretatie van de ernst van de benaderingsfout merken Browne & Cudeck (1992, p. 239) het volgende op. 'Practical experience has made us feel that a value of the RMSEA of about 0.05 or less would indicate a **close fit** of the model in relation to the degrees of freedom. ... We are also of the opinion that a value of about 0.08 or less for the RMSEA would indicate a reasonable error of approximation and would not want to employ a model with a RMSEA greater than 0.1.'

Met (14) en (21) volgt dat een puntschatting van de RMSEA wordt verkregen met de schatter

$$\hat{RMSEA} = (\hat{F}_0/d)^{1/2} = [\max(\hat{F}/d - 1/n, 0)]^{1/2} \quad . \tag{22}$$

Uit (11) en (21) volgt dat een schatting van het $100(1-\alpha)\%$-betrouwbaarheidsinterval voor de RMSEA, bij berekende schattingen van betrouwbaarheidsgrenzen voor de niet-centraliteitsparameter λ, te weten $\hat{\lambda}_L$ en $\hat{\lambda}_U$ [zie (19)], wordt gegeven door

$$((\hat{\lambda}_L/(nd))^{1/2}, (\hat{\lambda}_U/(nd))^{1/2}) \quad . \tag{23}$$

Browne & Cudeck (1992, p. 240) merken op dat het voordeel van het schatten van een betrouwbaarheidsinterval voor de RMSEA is, dat een bovengrens die groter is dan nul ons er nog eens aan herinnert dat het gepostuleerde model niet als een correct model kan worden beschouwd.

Een toets voor nabije passing

Zoals zijdelings al werd aangegeven maken Browne en Cudeck een onderscheid tussen de nulhypothese voor exacte passing (**'exact fit'**) en die voor nabije passing (**'close fit'**).

De nulhypothese voor exacte passing stelt dat er een parametervector $\boldsymbol{\theta}_0$ bestaat zodanig dat $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$; als dat het geval is zeggen we dat het model waar is. Voor deze hypothese wordt de toetsingsgrootheid $n\hat{F}$ [zie uitdrukking (1a)] beschouwd onder een centrale χ^2-verdeling met d vrijheidsgraden; zie voor details Boomsma (1994).

Omdat zij een dergelijke hypothese niet realistisch vinden geven Browne & Cudeck (1992, pp. 240-241) de voorkeur aan een toets voor nabije passing, dat wil zeggen aan een toets voor de nulhypothese dat RMSEA ≤ 0.05 tegen het alternatief dat RMSEA > 0.05. Ter toetsing van deze hypothese wordt eveneens de toetsingsgrootheid $n\hat{F}$ gebruikt, maar nu niet onder een centrale χ^2-verdeling met d vrijheidsgraden. In plaats daarvan wordt $G(x|\lambda,d)$ gebruikt, de verdelingsfunctie van de niet-centrale χ^2-verdeling met d vrijheidsgraden, terwijl uit (11) en (21) volgt dat in dit geval de niet-centraliteitsparame-

ter $\lambda = ndRMSEA^2$. Onder de nulhypothese RMSEA ≤ 0.05 volgt daar vervolgens uit dat de overschrijdingskans dat de stochastische variabele $n\underline{\hat{F}}$ een waarde aanneemt die groter is dan de gevonden waarde $n\hat{F}$, gelijk is aan

$$P\{n\underline{\hat{F}} > n\hat{F}\} = 1 - P\{n\underline{\hat{F}} \leq n\hat{F}\} = 1 - G(n\hat{F}|nd0.05^2,d) \quad . \qquad (24)$$

De nulhypothese wordt, bij een tevoren gekozen fout van de eerste soort α, verworpen als deze overschrijdingskans $P\{n\underline{\hat{F}} > n\hat{F}\} < \alpha$. De grootheid (24) wordt door LISREL 8 afgedrukt.

Verder merken Browne & Cudeck (1992, p. 241) nog op dat dezelfde hypothese niet zal worden verworpen op een niveau van $\alpha = 0.05$ als de benedengrens van het 90%-betrouwbaarheidsinterval voor RMSEA, gedefinieerd door (23) kleiner is dan 0.05.

Onderzoekers die een andere waarde bij de toets voor nabije passing willen gebruiken dan de gesuggereerde waarde van 0.05 kunnen de benodigde overschrijdingskans (24) op eenvoudige wijze bepalen met behulp van een electronische tabel, zoals bij voorbeeld StaTable (Metha & Patel, 1990).

Kruisvalidatie en de ECVI als schatting van de totale fout

We bespreken nu achtereenvolgens twee benaderingen die wellicht meer dan andere maten worden aangewend voor modelselectie, te weten het gebruik van kruisvalidatie en informatiecriteria. Vanwege de nauwe aansluiting van de eerste aanpak bij de stof die hiervoor is behandeld, bespreken we de kruisvalidatiebenadering eerst, daarna komen enkele informatiecriteria aan de orde. Hoewel de aanpak van beide benaderingen verschilt, zijn er, zoals we zullen zien, wel overeenkomsten aan te geven.

Op het belang van kruisvalidatie kan met name in het kader van exploratieve modelmodificatie worden gewezen (zie Camstra & Boomsma, 1992). Het grondidee van kruisvalidatie bij covariantiestructuuranalyse is dat we beschikken over twee onafhankelijke steekproeven uit dezelfde populatie: een **calibratiesteekproef** $\mathbf{S}_c$ en een **validatiesteekproef** $\mathbf{S}_v$. Het gepostuleerde model wordt geschat op basis van de calibratiesteekproef $\mathbf{S}_c$, hetgeen resulteert

in een modelgeïmpliceerde covariantiematrix $\hat{\Sigma}_c$. Vervolgens kijken we hoe goed het geschatte model past bij de onafhankelijke validatiesteekproef $\mathbf{S}_v$.

De **kruisvalidatie-index** ('cross-validation index') wordt door Cudeck & Browne (1983, p. 151) gedefinieerd als

$$CVI = F(\mathbf{S}_v;\hat{\Sigma}_c) \quad , \tag{25}$$

dus als de discrepantie tussen $\mathbf{S}_v$ en $\hat{\Sigma}_c$ bij een gegeven doelfunctie F. Hoe kleiner de waarde van de CVI des te meer vertrouwen heeft een onderzoeker in de houdbaarheid van zijn model in onafhankelijke, nieuwe steekproeven.

Browne & Cudeck (1992, p. 242) laten zien dat de conditionele verwachting van de CVI over validatiesteekproeven $\mathbf{S}_v$ bij een gegeven calibratiesteekproef, i.e. bij gegeven $\hat{\Sigma}_c$, ongeveer gelijk is aan

$$E_v\{CVI\} = E_v\{F(\mathbf{S}_v;\hat{\Sigma}_c)\} \approx F(\Sigma_0;\hat{\Sigma}_c) + k^*/n_v = F_t + k^*/n_v \quad , \tag{26}$$

waarbij $k^* = k(k+1)/2$ en $n_v + 1$ de omvang van de validatiesteekproef $\mathbf{S}_v$ is, terwijl E_v verwachting over validatiesteekproeven betekent. De CVI is dus een onzuivere schatter van de totale discrepantie F_t [zie (9)] bij een gegeven calibratiesteekproef. De mate van onzuiverheid hangt af van de omvang van de validatiesteekproef en van het aantal geobserveerde variabelen k.

Onderzoekers beschikken vaak niet over een grote steekproef. Omdat een aselecte splitsing van de totale steekproef in een calibratie- en een validatiedeel de beschikbare steekproefomvang met de helft reduceert, zou dat de totale discrepantie F_t voor de gegeven calibratiesteekproef doen toenemen. Om die reden hebben Browne & Cudeck (1989) een kruisvalidatie-index op basis van de enig beschikbare, totale steekproef ontwikkeld, waarbij de steekproefomvang dus niet hoeft te worden gehalveerd. Deze door hen voorgestelde methode is een enkele steekproefprocedure.

Die laatste aanpak resulteert in een schatter voor de niet-conditionele **verwachte waarde van de kruisvalidatie-index** ('expected cross-validation

index'), i.e. de verwachte waarde over calibratie- en validatiesteekproeven:

$$\begin{aligned} ECVI = E_c E_v(CVI) &\approx F[\Sigma_0;\tilde{\Sigma}_j] + (k^* + t)/n \\ &= F_0 + (k^* + t)/n = F_0 + k^*/n + t/n \quad , \end{aligned} \tag{27}$$

waarbij $k^* = k(k+1)/2$ en t het aantal te schatten modelparameters is.

Uit (27) en (12) volgt dat voor de EVCI, respectievelijk bij het onafhankelijkheidsmodel M_i, het gepostuleerde model M_j, en het verzadigde model M_s, de volgende puntschatters te gebruiken zijn:

$$\hat{ECVI}_i = F[\mathbf{S};\Sigma(\hat{\boldsymbol{\theta}}_i)] + 2t_i/n = \hat{F}_i + 2k/n \quad , \tag{28}$$

$$\hat{ECVI}_j = F[\mathbf{S};\Sigma(\hat{\boldsymbol{\theta}}_j)] + 2t_j/n = \hat{F}_j + 2t_j/n \quad , \tag{29}$$

$$\hat{ECVI}_s = 2t_s/n = k(k+1)/n \quad , \tag{30}$$

waarbij **S** de complete steekproefcovariantiematrix is met steekproefomvang N = n+1. De verwachte waarde van de schatter $\hat{ECVI}$ is bij benadering gelijk aan ECVI; zie Browne & Cudeck (1989, 1992) voor nadere details. Merk op dat de ECVI afhangt van het aantal te schatten parameters van het model en van de steekproefomvang. De ratio $2t_j/n$ is de penalty-functie van de $\hat{ECVI}_j$. Merk tevens op dat in de praktijk $\hat{ECVI}_s$ groter kan zijn dan $\hat{ECVI}_j$.

Uit (11) en (27) volgt dat een schatting van het $100(1-\alpha)$%-betrouwbaarheidsinterval voor de ECVI wordt gegeven door

$$((\hat{\lambda}_L + k^* + t)/n,\ (\hat{\lambda}_U + k^* + t)/n) \quad , \tag{31}$$

waarbij $\hat{\lambda}_L$ en $\hat{\lambda}_U$ berekende schattingen van betrouwbaarheidsgrenzen voor de niet-centraliteitsparameter λ zijn [zie (19)].

Voor de meest aannemelijke schattingsmethode hebben Browne & Cudeck (1989) verfijningen op de $\hat{\text{ECVI}}$ en op de schattingen van betrouwbaarheidsintervallen voor de ECVI aangebracht, die we hier niet behandelen, mede omdat ze door LISREL 8 niet worden berekend.

Browne & Cudeck (1992) leggen er de nadruk op dat het hun bij de enkele steekproefprocedure gaat om inspectie van de verwachte totale discrepantie $F_t = F[\Sigma_0;\hat{\Sigma}_j]$ over alle mogelijke calibratiesteekproeven. De ECVI is namelijk ook bij benadering een functie van F_t:

$$\text{ECVI} \approx E\{F[\Sigma_0;\hat{\Sigma}_j]\} + k^*/n = E\{F_t\} + k^*/n \quad . \tag{32}$$

Zie Browne & Cudeck (1992, p. 244) voor een afleiding. Hoe kleiner de schatting van de ECVI des te kleiner de totale fout.

Als we ervan uitgaan dat het aantal geobserveerde variabelen k constant is, is de rangorde van gepostuleerde modellen M_j, j=1,...,m, op basis van de $\hat{\text{ECVI}}$ dezelfde als die op basis van de verwachte totale discrepantie.

Als de meest aannemelijke schattingsprocedure (ML) wordt gebruikt blijkt $\hat{\text{ECVI}}$ in (28)-(30) een lineaire relatie te hebben met Akaike's informatiecriterium (AIC) en zal ze tot eenzelfde rangorde van modellen voeren als AIC. De herschaalde AIC [zie (36)] is gelijk aan $\hat{\text{ECVI}}_{ML}$. Daarmee kan de AIC dus ook worden gezien als een modelselectiecriterium dat het model kiest met de kleinste totale fout.

Informatiecriteria

Informatiecriteria zijn gebaseerd zijn op statistische informatietheorie (Kullback, 1959). Van alle informatiecriteria die zijn voorgesteld noemen we hier alleen AIC en CAIC, die beide door EQS en LISREL 8 worden uitgerekend, alsmede SIC (BIC). Voor een recent overzicht van informatiecriteria verwijzen we naar Haughton (1996) en Oud, Haughton & Jansen (1996). De Leeuw (1992) bespreekt het baanbrekende artikel van Akaike (1973). Hij geeft

tevens een korte inleiding tot informatie- en kruisvalidatiecriteria, overigens gevolgd door een herdruk van Akaike (1973).

Vooraf zij er ook op gewezen dat de definities van de verschillende informatiecriteria van auteur tot auteur, en van programma tot programma nogal eens verschillen. Meestal is dat een kwestie van schaling, die aan de selectievolgorde van modellen geen afbreuk doet. Een en ander kan wel zeer verwarrend zijn; zo gebruiken EQS en LISREL bij voorbeeld duidelijk verschillende definities voor de informatiecriteria AIC en CAIC.

Evenals bij de ECVI worden de criteria AIC en CAIC door EQS en LISREL 8 niet alleen berekend voor het gepostuleerde model M_j met t_j te schatten modelparameters en d_j vrijheidsgraden, maar ter vergelijking ook voor het onafhankelijkheidsmodel M_i met $t_i = k$ parameters en $d_i = k(k-1)/2$ vrijheidsgraden en het verzadigde model M_s met $t_s = k(k+1)/2$ parameters en geen vrijheidsgraden. Beide informatiematen gebruiken een penalty-functie.

Omdat de informatiematen vanuit de ML-schattingsprocedure zijn ontwikkeld zou in hun definitie expliciet de doelfunctie $F_{ML}[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)]$ moeten staan. Of liever nog, in overeenstemming met de oorspronkelijke formulering de te maximaliseren natuurlijke logarithme van de aannemelijkheidsfunctie, $\ln L[\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j);\mathbf{S}] = -(n/2)F_{ML}[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)]$. Door LISREL 8 worden de informatiematen echter ongeacht de schattingsmethode uitgerekend, vandaar dat in de formules voor de informatiecriteria hier ook een algemene doelfunctie wordt gebruikt.

a. Akaike's informatiecriterium: AIC

Akaike's (1973, 1974, 1987) informatiecriterium AIC is in LISREL 8, respectievelijk voor het onafhankelijkheidsmodel M_i, het gepostuleerde model M_j en het verzadigde model M_s, gedefinieerd als

$$AIC_i = (N-1)F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_i)] + 2t_i = n\hat{F}_i + 2t_i = n\hat{F}_i + 2k \quad , \tag{33}$$

$$AIC_j = (N-1)F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)] + 2t_j = n\hat{F}_j + 2t_j \quad , \tag{34}$$

$$AIC_s = 2t_s = k(k+1) \quad . \tag{35}$$

AIC corrigeert de χ^2-passingsmaat met een term die een functie is van het aantal te schatten parameters. Bij deze definitie van AIC selecteert een onderzoeker het model met de kleinste AIC-waarde.

Cudeck & Browne (1983, p. 154) gebruiken een herschaling van AIC_j, waar we eerder naar verwezen in relatie tot $\hat{ECVI}_j$ bij een ML-schattingsprocedure [zie uitdrukking (29)], door (34) te delen door n = N-1:

$$AIC_j^* = \hat{F}_j + 2t_j/n = \hat{ECVI}_j \quad . \tag{36}$$

McDonald & Marsh (1990, p. 251) zijn nogal kritisch over het gebruik van AIC bij modelselectie: 'In short, we suggest that the AIC cannot be used in practice as an objective means of model selection as the selection made, like that of the conventional chi-square test, is largely a function of the sample size.' Zoals Cudeck & Browne (1983) al eerder lieten zien groeit de complexiteit van het model dat middels de AIC wordt gekozen bij toenemende steekproefomvang. Akaike's infomatiecriterium neigt bij modelselectie tot overpassing ('overfitting'). Dit zou kunnen betekenen dat wanneer de steekproefomvang maar groot genoeg wordt genomen, altijd voor het verzadigde model wordt gekozen, en gegeven uitdrukking (36) geldt dit in even sterke mate voor de $\hat{ECVI}_j$. Monte-Carlo-onderzoek van Oud, Haughton & Jansen (1996) laat echter zien dat de kans op overpassing zich bij toenemende steekproefomvang lijkt te stabiliseren en grenswaarden ongelijk nul of één bereikt.

b. CAIC: een consistente AIC

Bozdogan's (1987) consistente versie van de AIC-grootheid (CAIC) kent een iets andere penalty-functie dan de AIC; de laatste term wordt vermenigvuldigd met een functie van de steekproefomvang. In LISREL 8 is de CAIC, respectievelijk voor het onafhankelijkheidsmodel M_i, het gepostuleerde model M_j en het verzadigde model M_s, gedefinieerd als

$$CAIC_i = n\hat{F}_i + [1 + \ln N]t_i = n\hat{F}_i + k\,[1 + \ln N] \quad , \tag{37}$$

$$CAIC_j = n\hat{F}_j + t_j[1 + \ln N] \quad , \tag{38}$$

$$CAIC_s = t_s[1 + \ln N] = k^*[1 + \ln N] \quad , \tag{39}$$

waarbij $k^* = k(k+1)/2$ en N de steekproefgrootte is. De penalty-functie hangt hier af van de steekproefgrootte N en van het aantal te schatten parameters. Het gevolg is dat de tendens tot het kiezen van complexe modellen (overpassing) hier geringer is dan bij gebruik van AIC. Deze tendens is in feite niet meer aanwezig bij CAIC: asymptotisch wordt dan het correcte model gekozen. Met consistentie wordt juist bedoeld dat er noch van overpassing noch van onderpassing sprake is als de steekproefomvang tot oneindig nadert.

Niettemin schrijven McDonald & Marsh (1990, p. 251) over het gebruik van de CAIC bij modelselectie het volgende: 'However, it would appear that such variants on the AIC cannot do more than increase the sample size at which the saturated model is selected. So again, the conclusion is that these selection criteria cannot be used in practical applications, however effective they may be shown to be in applications to artificial data generated by an actual restrictive model.'

Bozdogan (1987) noemt nog een tweede informatiecriterium CAICF, een consistente AIC met Fisher-informatie, die we niet bespreken.

c. Schwarz' informatiecriterium: SIC

Schwarz (1978) gebruikt een iets andere penalty-functie dan Akaike. Deze grootheid wordt niet door LISREL 8 en EQS uitgerekend. Respectievelijk voor het onafhankelijkheidsmodel M_i, het gepostuleerde model M_j en het verzadigde model M_s, is het informatiecriterium van Schwarz gedefinieerd als

$$SIC_i = (N-1)F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_i)] + t_i \ln N = n\hat{F}_i + k \ln N \quad , \tag{40}$$

$$SIC_j = (N-1)F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)] + t_j \ln N = n\hat{F}_j + t_j \ln N \quad , \tag{41}$$

$$SIC_s = t_s \ln N = k^* \ln N \quad , \tag{42}$$

waarbij $k^* = k(k+1)/2$ en N de steekproefomvang is.

Cudeck & Browne (1983, p. 154) gebruiken een herschaling van SIC_j, door (41) te delen door $n = N-1$:

$$SIC_j^* = \hat{F}_j + [t_j \ln N]/n \quad . \tag{43}$$

In de literatuur wordt Schwarz' informatiecriterium ook wel Bayes' informatiecriterium genoemd, afgekort als BIC, omdat de afleiding ervan gebaseerd is op asymptotisch gedrag van Bayesiaanse schatters. Zie bij voorbeeld Raftery (1993), die laat zien welke cruciale rol BIC speelt bij een Bayesiaanse aanpak van het modelselectieprobleem.

Globale passingsindexen

Naast de χ^2-toetsingsgrootheid, die eerder werd besproken, noemen we een viertal andere globale passingsindexen; Hu & Bentler (1995, p. 85) gebruiken de term absolute passingsindexen.

a. De 'goodness-of-fit'-index

Jöreskog & Sörbom (1981, p. I.40) introduceerden twee maten voor de relatieve mate waarin de (co)varianties in **S** voorspeld worden door $\hat{\boldsymbol{\Sigma}}$: GFI en AGFI. In 1985 kwamen Tanaka en Huba met een index voor algemene modelpassing voor de GLS-schattingsmethode, in een vorm die eerder al door Bentler (1983, p. 507) was geïntroduceerd. In navolging daarvan formuleerden Jöreskog & Sörbom (1988, p. 43) hun oorspronkelijke indexen voor ML- en ULS-schattingsmethoden op diezelfde algemene manier; Tanaka & Huba (1985) bewezen de gelijkheid van de oorspronkelijke definities van Jöreskog en Sörbom uit 1981 en de algemenere, equivalente definities, die hier worden gereproduceerd.

De passingsindex GFI is in algemene termen gedefinieerd als

$$GFI = 1 - (\mathbf{s} - \hat{\boldsymbol{\sigma}})' \mathbf{W}^{-1} (\mathbf{s} - \hat{\boldsymbol{\sigma}}) / \mathbf{s}' \mathbf{W}^{-1} \mathbf{s}$$

$$= 1 - F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)]/F[\mathbf{S};\boldsymbol{\Sigma}(\mathbf{0})] = 1 - \hat{F}_j/\hat{F}_u \quad , \tag{44}$$

waarbij $\mathbf{W}$ een passende gewichtenmatrix is, afhankelijk van de schattingsmethode (zie Bollen, 1989b, pp. 276-277) en $\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_j)$ bij gegeven model M_j. De teller van de grootheid 1 - GFI, $\hat{F}_j = F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)]$, is steeds het minimum van de doelfunctie bij gegeven parameterschattingen $\hat{\boldsymbol{\theta}}_j$, terwijl de noemer van 1 - GFI, $\hat{F}_u = F[\mathbf{S};\boldsymbol{\Sigma}(\mathbf{0})]$, de waarde van de doelfunctie is als er geen parameterschattingen worden gemaakt, i.e. als alle schattingen gelijk aan nul worden gesteld (in zekere zin de maximale waarde die de doelfunctie kan aannemen, aldus Saris & Stronkhorst, 1984, p. 229). De GFI is een getal dat de relatieve reproductie van de covarianties in $\mathbf{S}$ weergeeft middels de modelgeïmpliceerde covariantiematrix $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)$; de maat geeft aan hoeveel beter het gepostuleerde model M_j het doet ten opzichte van het ultieme nulmodel M_u. Ze geeft de winst aan van het schatten van een model ten opzichte van het in het geheel niet schatten van een model. De GFI kan een waarde aannemen in de range [0,1]; in die zin noemen we het een genormeerde index.

Mulaik et al. (1989, pp. 435-436) wijzen op verschillen tussen de genormeerde passingsindex (NFI), die we verderop bespreken, en de GFI, alsmede op de analogie tussen de GFI en de gekwadrateerde multipele correlatiecoëfficiënt, die algemeen te definiëren is als 1 - Var(E)/Var(T), waarbij Var(E) de foutenvariantie en Var(T) de totale variantie is (zie ook Tanaka, 1993).

Opgemerkt zij dat de steekproevenverdeling van GFI niet bekend is. We weten daarom niet hoe groot groot is, of hoe klein klein is. We hebben geen absolute standaard om de gevonden waarden mee te vergelijken. Het is een beschrijvende maat, waarvan we de statistische eigenschappen niet kennen. Maiti & Mukherjee (1990) laten evenwel zien dat voor een bepaalde klasse van modellen er een monotone relatie bestaat tussen GFI en de χ^2-passingsmaat.

b. De aangepaste 'goodness-of-fit'-index

Als de GFI wordt aangepast voor het aantal vrijheidsgraden van de twee modellen die tegen elkaar worden afgezet, krijgen we de aangepaste passings-

index. Het aantal vrijheidsgraden van het ultieme nulmodel M_u is gelijk aan $d_u = k(k+1)/2$, immers $t_u = 0$. Het aantal vrijheidsgraden van het gepostuleerde, geschatte model M_j is gelijk aan d_j. De aangepaste 'goodness-of-fit'-index, is nu gedefinieerd als

$$\begin{aligned} AGFI &= 1 - \{F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)]/d_j\}/\{F[\mathbf{S};\boldsymbol{\Sigma}(\mathbf{0})/d_u\}] \\ &= 1 - (d_u/d_j)(1 - GFI) \quad . \end{aligned} \tag{45}$$

Tanaka (1993, p. 20) beschouwt de aangepaste 'goodness-of-fit'-index als een voor onzuiverheid aangepaste GFI; vergelijk de aangepaste gekwadrateerde multipele correlatiecoëfficiënt (de 'adjusted or shrunken' R^2).

Jöreskog & Sörbom (1988, p. 43) suggereren dat de twee passingsindexen GFI en AGFI, in tegenstelling tot de χ^2-maat, niet afhankelijk zijn van de grootte van de steekproefomvang. Dat is echter een aanvechtbare uitspraak, waar ze, naar later blijkt, ook niet achter staan (Jöreskog & Sörbom, 1993a, p. 122). Er mag dan wel geen N in de formules (44) en (45) zitten, het is evenwel onmiskenbaar dat de teller en de noemer van de statistische grootheid $(\mathbf{s} - \hat{\boldsymbol{\sigma}})'\mathbf{W}^{-1}(\mathbf{s} - \hat{\boldsymbol{\sigma}})/\mathbf{s}'\mathbf{W}^{-1}\mathbf{s} = 1 - GFI$, wel degelijk van de steekproefomvang afhangen. Te verwachten valt dat elementen van de teller en de noemer veranderen bij toenemende steekproefomvang. Te verwachten valt dat de variabiliteit van de steekproefcovarianties zal afnemen met de steekproefomvang. Als het model waar is zal de teller bij toenemende steekproefomvang naar nul gaan, omdat de elementen van de vector $(\mathbf{s} - \hat{\boldsymbol{\sigma}})$ dan steeds kleiner worden. De onbekende steekproevenverdelingen van de maten GFI en AGFI hangen daarmee wel degelijk van de steekproefomvang N af. Anderson & Gerbing (1984) vonden in een simulatiestudie onder meer dat de gemiddelden van de steekproevenverdelingen van GFI en AGFI groter werden met een toename van de steekproefomvang. Zie ook Bollen (1990), die twee soorten effecten van steekproefomvang onderscheidt.

Was de GFI een genormeerde index, omdat zijn waarde tussen nul en één ligt, de AGFI is dat niet, omdat ze waarden kleiner dan nul kan aannemen. Mulaik et al. (1989, p. 440) merken daarom op: 'Thus the AGFI index does not

have the rational norm of a meaningful zero point as does the parsimonious-fit index of James et al. (1982).'

c. De voor spaarzaamheid gecorrigeerde 'goodness-of-fit'-index

James, Mulaik & Brett (1982, p. 155) definiëren de **spaarzaamheid** van een model M_j als

$$p_j = d_j/d_0 \quad , \tag{46}$$

dus als de ratio van het aantal vrijheidsgraden van het te schatten model M_j en het aantal vrijheidsgraden van het nulmodel dat als basismodel voor de passingsvergelijking wordt gekozen. Het nulmodel bij de GFI is model M_u, waarbij parameters op een waarde van nul worden gefixeeerd. De spaarzaamheid van een model M_j, zoals gedefinieerd in (46), wordt ook wel de spaarzaamheidsindex ('parsimony index'; Mulaik et al., 1989, p. 436) of de spaarzaamheidsratio ('parsimony ratio'; McDonald & Marsh, 1990, p. 249) genoemd.

Mulaik et al. (1989) stellen nu een iets andere correctie op de GFI voor dan de AGFI, hetgeen resulteert in wat zij de 'parsimony goodness-of-fit index' (PGFI) noemen. Het idee van Mulaik et al. is de GFI te vermenigvuldigen met de spaarzaamheidsratio, om daarmee spaarzaamheid in termen van aantal vrijheidsgraden en de mate van passing relatief een even groot gewicht te geven. De 'parsimony goodness-of-fit index' is derhalve gedefinieerd als

$$\begin{aligned} PGFI &= (d_j/d_u)\ \{1 - F[\mathbf{S};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)]/F[\mathbf{S};\boldsymbol{\Sigma}(\mathbf{0})]\} \\ &= (d_j/d_u)\ (1 - \hat{F}_j/\hat{F}_u) = p_j\ GFI \quad . \end{aligned} \tag{47}$$

d. De RMR als functie van de residuen

De wortel uit de gemiddelde gekwadrateerde residuen (de 'root mean square residuals'; RMR) is gedefinieerd als

$$RMR = \left[\sum_{i=1}^{k} \sum_{j=1}^{i} (s_{ij} - \hat{\sigma}_{ij})^2 / k^* \right]^{1/2} \quad , \tag{48}$$

waarbij $k^* = k(k+1)/2$. Het is dus een maat voor de grootte van 'het gemiddelde residu' en kan alleen geïnterpreteerd worden in relatie tot de relatieve grootte van de covarianties in de steekproefcovariantiematrix **S**. Dat kan het best als alle geobserveerde variabelen gestandaardiseerd zijn. Om die reden wordt door LISREL 8 de zogenaamde **gestandaardiseerde RMR** berekend. Hoewel de handboeken (Jöreskog & Sörbom, 1993a, 1993b) dat niet vermelden is deze grootheid de waarde van RMR, zoals gedefinieerd in (48), wanneer de geobserveerde variabelen gestandaardiseerd zouden zijn op een variantie van één, als er dus als het ware een correlatiematrix zou zijn geanalyseerd. De gestandaardiseerde RMR is een genormeerde maat: ze kan een waarde tussen nul en één aannemen. De ongestandaardiseerde RMR heeft een ondergrens van nul.

De RMR kan gebruikt worden om verschillende modellen bij dezelfde gegevens te vergelijken. Jöreskog & Sörbom (1988, p. 43) beweren dat de GFI dat ook kan en dat die maat ook kan worden gebruikt om modellen bij verschillende gegevens te vergelijken, hetgeen we betwijfelen (de GFI is weliswaar een genormeerde index, maar uit onderzoek blijkt bij voorbeeld dat de grootte van GFI van de steekproefomvang afhangt). Afgezien daarvan geldt dat de RMR en de GFI net als de globale χ^2-passingsmaat altijd maximaal tot overpassing leiden. Bij vergelijking van verschillende modellen voor dezelfde steekproefgegevens wordt altijd het minst restrictieve model gekozen. Alleen bij de AGFI is dit in mindere mate het geval.

Ten slotte merken de auteurs van het LISREL-handboek nog op dat de χ^2, GFI en RMR maten zijn voor de algemene passing van het model, het zijn samenvattende maten. Deze grootheden zeggen niets over de kwaliteit van het model in termen van andere interne en externe criteria. De globale passing van het model kan bij voorbeeld goed zijn, terwijl sommige verbanden afgaande op gekwadrateerde multipele correlatiecoëfficiënten erg slecht gemeten of bepaald zijn. Als het model niet goed bij de gegevens past geven deze globale maten ook nauwelijks een indicatie hoe dat komt, of in welk deel van het model de oorzaak van de slechte passing gevonden moet worden. Daarvoor kijken we naar

gedetailleerde informatie over de adequaatheid van het model, oftewel naar locale indicaties voor de adequaatheid van een model ('local assessment of fit').

Indexen voor het vergelijken van modellen: incrementele passingsindexen

De volgende zes passingsindexen zijn gebaseerd op de gedachte de passing van een model te beschouwen in relatie tot een zogenaamd referentiemodel, ook wel basismodel ('baseline model') of nulmodel ('null model') genoemd; de doelfunctie $F[\mathbf{S};\Sigma(\hat{\boldsymbol{\theta}})]$ die hierbij wordt gebruikt is willekeurig. Dit soort indexen worden algemeen 'incremental fit indices' genoemd. Zoals eerder opgemerkt wordt steeds verondersteld dat het basismodel M_b (of het nulmodel M_0) restrictiever is dan het gepostuleerde model M_j; onder M_j worden er meer parameters geschat dan onder M_b.

a. De genormeerde passingsindex NFI

De genormeerde passingsindex NFI ('normed fit index') van Bentler & Bonett (1980, p. 599), is in zijn eenvoudigste, maar meest algemene vorm (vergelijk echter Mulaik et al., 1989, p. 432) gedefinieerd als

$$NFI_b = (\hat{F}_b - \hat{F}_j)/\hat{F}_b = (\chi^2_b - \chi^2_j)/\chi^2_b = 1 - \hat{F}_j/\hat{F}_b = 1 - \chi^2_j/\chi^2_b \quad , \quad (49)$$

waarbij $\hat{F}_j = \chi^2_j/n$, $\hat{F}_b = \chi^2_b/n$ de waarden van de doelfuncties berekend bij $\Sigma(\hat{\boldsymbol{\theta}})$ voor respectievelijk M_j en een willekeurig basismodel M_b, met $n = N-1$. De NFI is een maat voor de proportionele vermindering van de waarde van de doelfunctie of de χ^2-waarde, als in plaats van een basismodel M_b gekozen wordt voor een minder restrictief model M_j. Ze representeert 'the proportion of the total lack of fit that has been reduced by the use of Model j.' (Mulaik et al., 1989, p. 433) In LISREL 8 wordt als basismodel standaard het onafhankelijkheidsmodel M_i genomen, dus

$$NFI_i = (\hat{F}_i - \hat{F}_j)/\hat{F}_i = (\chi^2_i - \chi^2_j)/\chi^2_i = 1 - \hat{F}_j/\hat{F}_i = 1 - \chi^2_j/\chi^2_i \quad . \quad (50)$$

Andere keuzes voor een basismodel kunnen evenwel meer voor de hand liggen (zie Cudeck & Browne, 1983; Sobel & Bohrnstedt, 1985).

Er is hier sprake van een genormeerde index, omdat zijn waarde altijd tussen nul en één ligt. McDonald & Marsh (1990, p. 250) wijzen er op dat deze Bentler-Bonett index in eindige steekproeven (zeg in steekproeven kleiner dan 800) een onderschatting geeft van de populatiewaarde.

b. De niet-genormeerde passingsindex NNFI

De niet-genormeerde passingsindex NNFI ('nonnormed fit index') van Bentler & Bonett (1980) is in de context van exploratieve factor-analyse eerder door Tucker & Lewis (1973) voorgesteld. Deze index houdt rekening met het aantal vrijheidsgraden en is, met het onafhankelijkheidsmodel als basismodel, gedefinieerd als

$$NNFI_i = (\hat{f}_i - \hat{f}_j)/(\hat{f}_i - 1) = (\chi_i^2/d_i - \chi_j^2/d_j)/(\chi_i^2/d_i - 1) \quad , \qquad (51)$$

waarbij $\hat{f}_i = n\hat{F}_i/d_i$ en $\hat{f}_j = n\hat{F}_j/d_j$ met $n = N-1$ (zie ook Bollen, 1986). Er is sprake van een niet-genormeerde index, omdat zijn waarde niet altijd tussen nul en één ligt.

Over het verschil tussen de NNFI ('the Tucker-Lewis index') en de NFI ('the Bentler-Bonett index') merken McDonald & Marsh (1990, p. 250) op: 'Rather the important difference is that the former is an (unbiased) estimator of a quantity that incorporates the parsimony ratio, and the latter is a (biased) estimator of a quantity that does not incorporate the parsimony ratio.'

c. De voor spaarzaamheid gecorrigeerde passingsindex PNFI

De voor spaarzaamheid gecorrigeerde passingsindex PNFI ('parsimony normed fit index') van James, Mulaik & Brett (1982, p. 155), die het aantal vrijheidsgraden als maat voor spaarzaamheid verdisconteerd, is gedefinieerd als

$$PNFI_i = (d_j/d_i)(1 - \hat{F}_j/\hat{F}_i) = (d_j/d_i)\, NFI_i \quad . \qquad (52)$$

Merk op dat deze maat veel overeenkomst vertoont met de PGFI [zie vergelijking (47)]. Het nulmodel bij PNFI is hier echter het onafhankelijkheidsmodel M_i, terwijl dat voor de PGFI het ultieme nulmodel M_u is.

Mulaik et al. (1989, p. 440) introduceerden nog een tweede soort PNFI.

d. De vergelijkende passingsindexen CFI en FI

Om onder- en overschattingsproblemen van de NFI en de NNFI tegen te gaan heeft Bentler (1990a) nieuwe genormeerde (CFI) en niet-genormeerde (FI) indexen geïntroduceerd.

De genormeerde vergelijkende passingsindex CFI ['normed comparative fit index'; Bentler, 1990a, formule 13] is gedefinieerd als

$$CFI_i = 1 - \hat{\tau}_j / \hat{\tau}_i \quad , \tag{53}$$

waarbij voor het gepostuleerde model $\hat{\tau}_j = \max[(n\hat{F}_j - d_j), 0] = \max[\hat{\lambda}_j, 0]$ en voor het onafhankelijkheidsmodel $\hat{\tau}_i = \max[(n\hat{F}_i - d_i), (n\hat{F}_j - d_j), 0] = \max[\hat{\lambda}_i, \hat{\lambda}_j, 0]$.

De niet-genormeerde vergelijkende passingsindex FI ['nonnormed comparative fit index'; Bentler, 1990a, formule (12)], die gelijk is aan de relatieve niet-centraliteitsindex (RNI) van McDonald & Marsh (1990, p. 250, formule 27), is gedefinieerd als

$$\begin{aligned} FI_i &= 1 - (n\hat{F}_j - d_j)/(n\hat{F}_i - d_i) = 1 - (\chi^2_j - d_j)/(\chi^2_i - d_i) \\ &= 1 - \hat{\lambda}_j / \hat{\lambda}_i \quad . \end{aligned} \tag{54}$$

Deze index wordt niet door LISREL 8 uitgerekend.

Bentler (1990a) bespreekt deze indexen in termen van het vergelijken van niet-centraliteitsparameters; $n\hat{F}_j - d_j$ is immers een schatter voor λ_j [vergelijk (15)]. De algemene bedoeling bij het gebruik van deze indexen is een schatting te krijgen van de populatiegrootheid $\delta = 1 - \lambda_j / \lambda_b$, waarbij

$\lambda_b \geq \lambda_j$ de niet-centraliteitsparameter van een basismodel M_b is.

e. De incrementele passingsindex IFI

De incrementele passingsindex IFI ('incremental fit index') van Bollen (1989b, p. 271, formule 7.53; zie ook Bollen, 1989a), die gelijk is aan de index c_k bij McDonald & Marsh (1990, p. 250, formule 24), is een modificatie van de NFI. Het effect van deze modificatie is dat de verwachting van IFI minder afhankelijk is van de steekproefomvang N, terwijl ook rekening wordt gehouden met het aantal vrijheidsgraden van M_j. Met het onafhankelijkheids-model als basismodel is deze index gedefinieerd als

$$IFI_i = (\hat{F}_i - \hat{F}_j)/(\hat{F}_i - d_j/n) = (\chi_i^2 - \chi_j^2)/(\chi_i^2 - d_j) \quad . \tag{55}$$

f. De relatieve passingsindex RFI

De relatieve passingsindex RFI ('relative fit index') van Bollen (1986; 1989b, p. 272, formule 7.54; zie ook Bollen, 1989a) is ook een modificatie van de NFI. De bedoeling van deze modificatie is het gepostuleerde model M_j af te zetten tegen een basismodel M_b, en die vergelijking per vrijheidsgraad van beide modellen M_j en M_b te doen. Met het onafhankelijkheidsmodel als basismodel is deze index gedefinieerd als

$$RFI_i = (\hat{f}_i - \hat{f}_j)/\hat{f}_i = (\chi_i^2/d_i - \chi_j^2/d_j)/(\chi_i^2/d_i) \quad . \tag{56}$$

De critische N-grootheid CN

De critische N-grootheid (CN), die door Hoelter (1983) werd voorgesteld, is bedoeld een indicatie te geven van de steekproefomvang N, die nodig zou zijn om de χ^2-waarde, die correspondeert met model M_j, net significant te doen zijn op een significantieniveau ter grootte van α. Hij geeft de steekproef-omvang aan waarbij de waarde van de doelfunctie F_j tot verwerping van de

nulhypothese $\Sigma = \Sigma(\theta_j)$ zou leiden, en is gedefinieerd als

$$CN = 1 + \chi^2_{1-\alpha}/\hat{F}_j \quad , \tag{57}$$

waarbij $\chi^2_{1-\alpha}$ het $(1-\alpha)$-percentielpunt is onder de centrale χ^2-verdeling met d_j vrijheidsgraden. Merk op dat LISREL 8 bij de berekening van de grootheid CN de waarde $\alpha = 0.01$ gebruikt.

Zie Bollen & Liang (1988) en Bollen (1989b, p. 277f.) voor een kritische bespreking van deze statistische grootheid: 'CN may lead to an overly pessimistic assessment of fit for small samples.' (Bollen, 1989b, p. 278)

Hoe te kiezen?

Voor een overzicht van de vele in omloop zijnde indexen en van empirisch onderzoek naar hun gedrag verwijzen we naar De Gooijer & Koopman (1988), Marsh, Balla & McDonald (1988), Bollen (1989b, pp. 269-281), La Du & Tanaka (1989), Mulaik et al. (1989), Bentler (1990a, 1990b), McDonald & Marsh (1990), Bandalos (1993), Hu, Bentler & Kano (1992), Bentler & Chou (1993), Camstra (1993), Gerbing & Anderson (1993), Tanaka (1993), Williams & Holahan (1994), Ding, Velicer & Harlow (1995), en Oud, Haughton & Jansen (1996). De rol van passingsindexen bij het evalueren van modellen wordt ook besproken door Kaplan (1990) en door Hu & Bentler (1995). Een veelbelovende kruisvalidatieprocedure is die van De Gooijer & Koopman (1988), die ook door Camstra & Boomsma (1992) wordt besproken; uit vergelijkend onderzoek (Camstra, 1983) blijkt deze procedure tot goede resultaten te leiden.

Op een tweetal aspecten van indexkeuze willen we hier ten slotte wijzen. Het eerste aspect is dat Marsh et al. (1988) twee soorten incrementele passingsmaten onderscheiden.

Type 1 is van de vorm

$$IFI1(F) = (\hat{F}_b - \hat{F}_j)/\hat{F}_b \quad , \tag{58}$$

waarbij F een gekozen passingsmaat, zoals een doelfunctie F, een χ^2-waarde, de ratio χ^2/d, of RMR; $\hat{F}_b$ is de waarde van die passingsmaat voor een gekozen basismodel M_b en $\hat{F}_j$ de waarde van die maat voor het gepostuleerde model M_j (zie ook Mulaik et al., 1989, p. 434).

Type 2 is van de vorm

$$IFI2(F) = (\hat{F}_b - \hat{F}_j)/(\hat{F}_b - E(\hat{F}_j|M_j \text{ is waar})) \quad . \tag{59}$$

Anders dan bij indexen van het Type 1 wordt bij indexen van het Type 2 dus rekening gehouden met de verwachte waarde van $\hat{F}_j$, onder het gegeven dat model M_j waar is. De NFI is een voorbeeld van Type 1, de IFI een voorbeeld van Type 2. Marsh et al. (1988) vonden dat indexen van het Type 1 gemiddeld, bij kleine steekproeven ($N < 200$), de asymptotische waarde van de betreffende index significant onderschatten. Indexen van het Type 2 onderschatten de asymptotische waarde in kleine steekproeven in veel mindere mate dan indexen van het Type 1. Vandaar dat ze in hun aanbevelingen een sterke voorkeur uitspreken voor indexen van het Type 2.

Hu & Bentler (1995) noemen incrementele indexen waarbij de niet-centraliteitsparameter centraal staat, zoals de vergelijkende passingsindexen CFI en FI [zie (53) en (54)], passingsindexen van het Type 3.

Het tweede aspect dat we willen noemen is dat Tanaka (1993) een categoriseringsschema voor passingsmaten heeft opgesteld, waarbij een index wel of niet tot een zestal categorieën (door hem als dimensies gekarakteriseerd) behoort. We geven de zes dimensies kort weer (o.c., p. 16, Tabel 2.1).

1. Populatie- versus steekproefgeoriënteerde maten; het schatten van een populatiegrootheid versus het beschrijven van de modelpassing op basis van de steekproefgegevens. Voorbeeld: IFI versus PGFI.
2. Eenvoud versus complexiteit; passingsmaten met een voorkeur voor eenvoudige modellen hanteren een correctie- of penalty-functie om modellen met veel parameters te corrigeren. Voorbeeld: AIC versus GFI.
3. Genormeerde versus niet-genormeerde maten; genormeerde indexen liggen in een vaste range, meestal [0,1]. Voorbeeld: NFI versus NNFI.
4. Absolute versus relatieve passingsmaten; relatieve maten gebruiken ter

evaluatie van de modelpassing een vergelijking met een of ander basis-model. Voorbeeld: AIC versus CFI.

5. Passingsmaten die conditioneel invariant zijn onder de keuze van de schattingsmethode versus maten die dat niet zijn; in het eerste geval zijn de waarden van de passingsmaten ongeveer gelijk als er verschillende schattingsmethoden bij dezelfde steekproefgegevens worden gebruikt (zie ook La Du & Tanaka, 1989). Voorbeeld: NNFI versus NFI.
6. Maten die onafhankelijk zijn van de steekproefomvang versus maten die daarvan wel afhankelijk zijn; in het eerste geval hangt de waarde van de index af van de steekproefomvang N (zie ook Bollen, 1990). Voorbeeld: IFI versus AIC.

   Overigens beschouwt Tanaka alleen effecten van N op het gemiddelde van de steekproevenverdeling van een passingsmaat. Wel constateert hij het gegeven dat anderen menen dat de steekproefomvang altijd een effect heeft op de steekproefvariabiliteit van **S** en daarmee mogelijk op elk schattingsresultaat.

Het is voor de gebruiker van structurele modellen interessant na te lezen op basis van welke argumenten Tanaka, gegeven deze zes dimensies, richtlijnen geeft voor een selectie van passingsmaten en welke controverses daarbij kunnen optreden.

Een praktisch voorbeeld

Om een eerste indruk te krijgen van de LISREL-uitvoer geven we een overzicht van de behandelde maten en indexen aan de hand van voorbeeld 6.4 uit de LISREL-handleiding (Jöreskog & Sörbom, 1988, pp. 169-177; 1993b, pp. 201-208). In dit voorbeeld over de stabiliteit van vervreemding (Wheaton, Muthén, Alwin & Summers, 1977) worden vier modellen vergeleken. Model A is een spaarzaam model met vier geobserveerde variabelen. Model B, C en D, elk met twee extra geobserveerde variabelen, zijn in toenemende mate complex. De analyses zijn telkens gebaseerd op een steekproefomvang van N = 932. Tabel 1 bevat de vier invoerfiles, waarbij de SIMPLIS-commandotaal wordt gebruikt.

De verkregen resultaten wat betreft de passingsmaten zijn samengevat in Tabel 2. De volgorde van de maten in die tabel is voor de overzichtelijkheid veranderd ten opzichte van de LISREL-uitvoer. De notatie -- in de tabel

Tabel 1. SIMPLIS-invoerfiles van vier modellen voor de stabiliteit van vervreemding.

Stability of Alienation - Model A: No Ses, and Uncorrelated Errors

```
Observed Variables
   ANOMIA67   POWERL67   ANOMIA71   POWERL71   EDUC   SEI
Covariance Matrix
   11.834
    6.947      9.364
    6.819      5.091     12.532
    4.783      5.028      7.495      9.986
   -3.839     -3.889     -3.841     -3.625      9.610
   -2.190     -1.883     -2.175     -1.878      3.552    4.503
Sample Size 932
Latent Variables  Alien67  Alien71
Paths
   Alien67  ->  ANOMIA67  POWERL67
   Alien71  ->  ANOMIA71  POWERL71
   Alien67  ->  Alien71
End of Problem
```

Stability of Alienation - Model B: Uncorrelated Errors

```
Observed Variables
   ANOMIA67   POWERL67   ANOMIA71   POWERL71   EDUC   SEI
Covariance Matrix from File EX06.COV
Sample Size 932
Latent Variables  Alien67  Alien71  Ses
Paths
   Alien67  ->  ANOMIA67  POWERL67
   Alien71  ->  ANOMIA71  POWERL71
   Ses      ->  EDUC  SEI
   Alien67  ->  Alien71
   Ses      ->  Alien67  Alien71
End of Problem
```

Tabel 1 (vervolg). SIMPLIS-invoerfiles van vier modellen voor de stabiliteit van vervreemding.

Stability of Alienation - Model C: Correlated Errors of ANOMIA only

```
Observed Variables
   ANOMIA67   POWERL67   ANOMIA71   POWERL71   EDUC  SEI
Covariance Matrix from File EX06.COV
Sample Size 932
Latent Variables  Alien67  Alien71  Ses
Paths
   Alien67  ->  ANOMIA67  POWERL67
   Alien71  ->  ANOMIA71  POWERL71
   Ses      ->  EDUC  SEI
   Alien67  ->  Alien71
   Ses      ->  Alien67  Alien71
Let the Errors of ANOMIA67 and ANOMIA71 Correlate
End of Problem
```

Stability of Alienation - Model D: Correlated Errors of ANOMIA and POWERL

```
Observed Variables
   ANOMIA67   POWERL67   ANOMIA71   POWERL71   EDUC  SEI
Covariance Matrix from File EX06.COV
Sample Size 932
Latent Variables  Alien67  Alien71  Ses
Paths
   Alien67 ->  ANOMIA67  POWERL67
   Alien71 ->  ANOMIA71  POWERL71
   Ses     ->  EDUC  SEI
   Alien67 ->  Alien71
   Ses     ->  Alien67  Alien71
Let the Errors of ANOMIA67 and ANOMIA71 Correlate
Let the Errors of POWERL67 and POWERL71 Correlate
End of Problem
```

Tabel 2. Passingsmaten bij vier modellen voor de stabiliteit van vervreemding (N = 932).[1]

	Model A	Model B	Model C	Model D
Postulated model M_j				
Observed variables k	4	6	6	6
$k^* = k(k+1)/2$	10	21	21	21
Estimated parameters t_j	9	15	16	17
Test for exact fit				
χ^2 statistic $n\hat{F}$	61.11	71.47	6.33	4.73
Degrees of freedom d_j	1	6	5	4
P-value: $P(n\underline{\hat{F}} > n\hat{F})$	0.0	0.0	0.27	0.32
Mimimum fit function value $\hat{F}$	0.066	0.077	0.0068	0.0051
Independence model M_i				
χ^2 independence model M_i	1563.94	2131.40	2131.40	2131.40
Degrees of freedom d_i	6	15	15	15
Non-centrality				
Non-centrality parameter (NCP)	60.11	65.47	1.33	0.73
90% conf. interval for NCP	--	--	(0,12.07)	(0,10.53)
Error of approximation				
Population discrepancy (F0)	0.065	0.070	0.0014	0.00079
90% conf. interval for F0	--	--	(0,.013)	(0,.011)
Test for close fit				
RMSEA	0.25	0.11	0.017	0.014
90% conf. interval for RMSEA	--	--	(0,.051)	(0,.053)
P-value: $P(n\underline{\hat{F}} > n\hat{F})$	0.00000044	0.0000095	0.94	0.93
Expected cross-validation index				
ECVI independence model M_i	1.69	2.30	2.30	2.30
ECVI model M_j	0.085	0.11	0.041	0.042
90% conf. interval for ECVI	--	--	(.040,.053)	(.041,.052)
ECVI saturated model M_s	0.021	0.045	0.045	0.045

Tabel 2 (vervolg). Passingsmaten bij vier modellen voor de stabiliteit van vervreemding (N = 932).[1]

	Model A	Model B	Model C	Model D
Information criteria				
AIC independence model M_i	1571.94	2143.40	2143.40	2143.40
AIC model M_j	79.11	101.47	38.33	38.73
AIC saturated model M_s	20.00	42.00	42.00	42.00
CAIC independence model M_i	1595.29	2178.42	2178.42	2178.42
CAIC model M_j	131.64	189.03	131.73	137.97
CAIC saturated model M_s	78.37	164.58	164.58	164.58
Global fit measures				
Goodness-of-fit index (GFI)	0.97	0.98	1.00	1.00
Adjusted GFI (AGFI)	0.69	0.91	0.99	0.99
Parsimony GFI (PGFI)	0.097	0.28	0.24	0.19
Root mean square residual (RMR)	0.29	0.23	0.082	0.057
Standardized RMR (SRMR)	0.027	0.021	0.011	0.0074
Incremental fit indices				
Normed fit index (NFI)	0.96	0.97	1.00	1.00
Non-normed fit index (NNFI)	0.77	0.92	1.00	1.00
Parsimony NFI (PNFI)	0.16	0.39	0.33	0.27
Comparative fit index (CFI)	0.96	0.97	1.00	1.00
Incremental fit index (IFI)	0.96	0.97	1.00	1.00
Relative fit index (RFI)	0.77	0.92	0.99	0.99
Critical sample size				
Critical N | α = 0.01 (CN)	102.09	219.99	2218.49	2611.90

[1] -- betrouwbaarheidsintervallen zijn met onvoldoende precisie te schatten.

Tabel 3. Overzicht van LISREL-uitvoer en -definities

	Formule	J & S
CHI-SQUARE WITH d_j DEGREES OF FREEDOM	(1b)	(4.1)
P-VALUE FOR TEST OF EXACT FIT ($n\hat{F} > n\hat{F}$)	(5)	
ESTIMATED NON-CENTRALITY PARAMETER (NCP)	(15)	(4.3)
90 PERCENT CONFIDENCE INTERVAL FOR NCP	(19)	
MINIMUM FIT FUNCTION VALUE	(2)	
POPULATION DISCREPANCY FUNCTION VALUE (F0)	(14)	(4.13)
90 PERCENT CONFIDENCE INTERVAL FOR F0	(20)	(4.14)
ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA)	(22)	(4.15)
90 PERCENT CONFIDENCE INTERVAL FOR RMSEA	(23)	(4.16)
P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05)	(24)	(4.17)
EXPECTED CROSS-VALIDATION INDEX (ECVI)	(29)	(4.9)
90 PERCENT CONFIDENCE INTERVAL FOR ECVI	(31)	(4.10)
ECVI FOR SATURATED MODEL	(30)	
ECVI FOR INDEPENDENCE MODEL	(28)	
CHI-SQUARE FOR INDEPENDENCE MODEL WITH d_i DF	(3)	
INDEPENDENCE AIC	(33)	
MODEL AIC	(34)	(4.7)
SATURATED AIC	(35)	
INDEPENDENCE CAIC	(37)	
MODEL CAIC	(38)	(4.8)
SATURATED CAIC	(39)	
(STANDARDIZED) ROOT MEAN SQUARE RESIDUAL (RMR)	(48)	
GOODNESS OF FIT INDEX (GFI)	(44)	(4.11)
ADJUSTED GOODNESS OF FIT INDEX (AGFI)	(45)	(4.12)
PARSIMONY GOODNESS OF FIT INDEX (PGFI)	(47)	(4.24)
NORMED FIT INDEX (NFI)	(50)	(4.18)
NON-NORMED FIT INDEX (NNFI)	(51)	(4.20)
PARSIMONY NORMED FIT INDEX (PNFI)	(52)	(4.19)
COMPARATIVE FIT INDEX (CFI)	(53)	(4.21)
INCREMENTAL FIT INDEX (IFI)	(55)	(4.22)
RELATIVE FIT INDEX (RFI)	(56)	(4.23)
CRITICAL N (CN)	(57)	(4.25)

betekent dat de betreffende betrouwbaarheidsintervallen niet zijn berekend, omdat de P-waarde van de χ^2-toetsingsgrootheid $n\hat{F}$ te klein is; in feite kan LISREL 8 de betrouwbaarheidsgrenzen voor de niet-centraliteitsparameter [zie (18)] dan met onvoldoende precisie uitrekenen.

In Tabel 3 ten slotte staat een overzicht van de LISREL 8-uitvoer met verwijzing naar de formules in dit artikel en naar die in Jöreskog & Sörbom (1993a). Wat we missen in LISREL 8 bij de meest aannemelijke schattingsmethode is de herschaalde χ^2-grootheid van Satorra & Bentler (1988a, 1988b, 1994). Uit robuustheidsonderzoek van Hu, Bentler & Kano (1992) blijkt dat deze herschaalde grootheid robuust is bij de analyse van variabelen die niet een multivariate normale verdeling hebben en bij het gebruik van kleine steekproeven. Een tweede belangrijke aanvulling op LISREL-uitvoer zou, gegeven de resultaten van Oud, Haughton & Jansen (1996), de berekening van SIC (en liefst een van zijn verfijningen; o.c., formule 8) zijn.

Referenties

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Czaki (Eds.), Proceedings of the 2nd International Symposium on Information Theory (pp. 267-281). Budapest: Akademiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723.

Akaike, H. (1987). Factor analysis and AIC. Psychometrika, 52, 317-332.

Arbuckle, J. (1993). AMOS. Analysis of moment structure. Philadelphia, PA: Temple University, Department of Psychology.

Anderson, J.C., & Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions, and goodness of fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49, 155-173.

Bandalos, D.L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. Multivariate Behavioral Research, 28, 351-374.

Bentler, P.M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. Psychometrika, 48, 493-517.

Bentler, P.M. (1989). EQS: Structural equations program manual (Version 3.0). Los Angeles, CA: BMDP Statistical Software.

Bentler, P.M. (1990a). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bentler, P.M. (1990b). Fit indexes, Lagrange multipliers, constraint changes and incomplete data in structural models. Multivariate Behavioral Research, 25, 163-172.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bentler, P.M., & Chou, C.-P. (1993). Some new covariance structure model improvement statistics. In K.A. Bollen & J.S. Long (Eds.), Testing structural equation models (pp. 235-255). Newbury Park, CA: Sage.

Bentler, P.M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. Psychological Bulletin, 106, 315-317.

Bollen, K.A. (1986). Sample size and Bentler and Bonett's nonnormed fit index. Psychometrika, 51, 375-377.

Bollen, K.A. (1989a). A new incremental fit index for general structural equation models. Sociological Methods & Research, 17, 303-316.

Bollen, K.A. (1989b). Structural equations with latent variables. New York: Wiley.

Bollen, K.A. (1990). Overall fit in covariance structure models: Two types of sample size effects. Psychological Bulletin, 107, 256-259.

Bollen, K.A., & Liang, J. (1988). Some properties of Hoelter's CN. Sociological Methods & Research, 16, 492-503.

Boomsma, A. (1994). Analyse van covariantiestructuren. De adequaatheid van het model 1. Ongepubliceerd manuscript, Rijksuniversiteit Groningen, Vakgroep Statistiek en Meettheorie.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52, 345-370.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Browne, M.W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. Multivariate Behavioral Research, 24, 445-455.

Browne, M.W., & Cudeck, R. (1992). Alternative ways of assessing model fit. Sociological Methods & Research, 21, 230-258.

Browne, M.W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. British Journal of Mathematical and Statistical Psychology, 41, 193-208.

Camstra, A. (1993). The use of cross-validation techniques for model selection in covariance structure analysis. In R. Steyer, K.F. Wender & K.F. Widaman (Eds.), Psychometric methodology: Proceedings of the 7th European Meeting of the Psychometric Society in Trier (pp. 85-89). Stuttgart: Fischer.

Camstra, A., & Boomsma, A. (1992). Cross-validation in regression and covariance structure analysis: An overview. Sociological Methods & Research, 21, 89-115.

Cudeck, R., & Browne, M.W. (1983). Cross-validation of covariance structures. Multivariate Behavioral Research, 18, 147-167.

Cudeck, R., & Henly, S.J. (1991). Model selection in covariance structure analysis and the "problem" of sample size: A clarification. Psychological Bulletin, 109, 512-519.

De Gooijer, J.G., & Koopman, S.J. (1988). Cross-validation criteria and the analysis of covariance structures. In M.G.H. Jansen & W.H. van Schuur (Eds.), The many faces of multivariate analysis: Proceedings of the SMABS-88 Conference (Vol. II, pp. 296-311). Groningen: RION.

De Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. Journal of Econometrics, 22, 113-137.

De Leeuw, J. (1988). Model selection in multinomial experiments. In T.K. Dijkstra (Ed.), On model uncertainty and its statistical implications (pp. 118-138). Berlin: Springer.

De Leeuw, J. (1990). Data modeling and theory construction. In J. Hox & A. de Jonge-Gierveld (Eds.), Operationalization and research strategy (pp. 229-246). Amsterdam: Swets & Zeitlinger.

De Leeuw, J. (1992). Introduction to Akaike (1973) Information theory and an extension of the maximum likelihood principle. In S. Kotz & N.L. Johnson (Eds.), Breakthroughs in statistics (Vol. I: Foundations and basic theo-

ry; pp. 599-609). New York: Springer.

Ding, L., Velicer, W.F., & Harlow, L.L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. Structural Equation Modeling, 2, 119-143.

Gerbing, D.W., & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen & J.S. Long (Eds.), Testing structural equation models (pp. 40-65). Newbury Park, CA: Sage.

Haughton, D.M.A. (1996). Information criteria for model selection. Kwantitatieve Methoden (this issue).

Hoelter, J.W. (1983). The analysis of covariance structures: Goodness-of-fit indices. Sociological Methods & Research, 11, 325-344.

Hu, L.-T., & Bentler, P.M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 76-99). Thousand Oaks, CA: Sage.

Hu, L.-T., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin, 112, 351-362.

James, L.R., Mulaik, A.M., & Brett, J.M. (1982). Causal analysis: Assumptions, models, and data. Newbury Park, CA: Sage.

Jöreskog, K.G., & Sörbom, D. (1981). LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods (Research Report 81-8). University of Uppsala, Department of Statistics.

Jöreskog, K.G., & Sörbom, D. (1988). LISREL 7: A guide to the program and applications. Chicago, IL: SPSS.

Jöreskog, K.G., & Sörbom, D. (1993a). LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago, IL: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1993b). LISREL 8: User's reference guide. Chicago, IL: Scientific Software International.

Kaplan, D.W. (1990). Evaluating and modifying covariance structure models: A review and a recommendation (with discussion). Multivariate Behavioral Research, 25, 137-155.

Kuipers, T.A.F. (1992). Naive and refined truth approximation. Synthese, 93, 299-241.

Kullback, S. (1959). Information theory and statistics. New York: Wiley.

La Du, T.J., & Tanaka, J.S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit asssessments in structural equation models. Journal of Applied Psychology, 74, 625-635.

Linhart, H., & Zucchini, W. (1986). Model selection. New York: Wiley.

Maiti, S.S., & Mukherjee, B.N. (1990). A note on distributional properties of the Jöreskog-Sörbom fit indices. Psychometrika, 55, 721-726.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

McDonald, R.P. (1989). An index of goodness of fit based on noncentrality. Journal of Classification, 6, 97-103.

McDonald, R.P., & Marsh, H.W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. Psychological Bulletin, 107, 247-255.

Metha, C.R., & Patel, N.R. (1990). Electronic tables for statisticians and engineers (Version 1.0). Cambridge, MA: Cytel Software.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S., & Stillwell, C.D. (1989). An evaluation of goodness of fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Oud, J.H.L., Haughton, D.M.A., & Jansen, R.A.R.G. (1996). Information and other criteria in structural equation model selection. Kwantitatieve Methoden (this issue).

Raftery, A.E. (1993). Bayesian model selection in structural equation models. In K.A. Bollen & J.S. Long (Eds.), Testing structural equation models (pp. 163-180). Newbury Park, CA: Sage.

Saris, W.E., & Stronkhorst, H. (1984). Causal modelling in nonexperimental research: An introduction to the LISREL approach. Amsterdam: Sociometric Research Foundation.

Satorra, A., & Bentler, P.M. (1988a). Scaling corrections for chi-square statistics in covariance structure analysis. Proceedings of the Business and Economic Statistics Section of the ASA (pp. 308-313). Alexandria, VA: The American Statistical Association.

Satorra, A., & Bentler, P.M. (1988b). Scaling corrections for statistics in covariance structure analysis (UCLA Statistics Series No. 2). Los Angeles, CA: University of California, Department of Psychology.

Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), Latent variable analysis: Applications for developmental research (pp. 399-419). Thousand Oaks, CA: Sage.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461-464.

Shapiro, A. (1985). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. South African Statistical Journal, 19, 73-81.

Sobel, M.E., & Bohrnstedt, G.W. (1985). Use of null models in evaluating the fit of covariance structure models. In N.B. Tuba (Ed.), Sociological Methodology 1985 (pp. 152-178). San Francisco: Jossey-Bass.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research, 25, 173-180.

Steiger, J.H., & Lind, J.C. (1980). Statistically-based tests for the number of common factors. Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City, IA.

Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In K.A. Bollen & J.S. Long (Eds.), Testing structural equation models (pp. 10-39). Newbury Park, CA: Sage.

Tanaka, J.S., & Huba, G.J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. British Journal of Mathematical and Statistical Psychology, 38, 197-201.

Tucker, L., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Wheaton, B., Muthén, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models. In D.R. Heise (Ed.), Sociological Methodology 1977 (pp. 84-136). San Francisco: Jossey-Bass.

Williams, L.J., & Holahan, P.J. (1994). Parsimony-based fit indices for multiple indicator models: Do they work? Structural Equation Modeling, 1, 161-189.

Ontvangen: 1 juli 1995
Geaccepteerd: 2 maart 1996