KM 16(1984)
pag 21 - 39

ROBUUSTHEIDSONDERZOEK VAN LISREL

A. Boomsma

Samenvatting

Er wordt een beknopt overzicht gegeven van een onderzoeksaanpak naar aanleiding van een tweetal robuustheidsvragen die bij de analyse van covariantiestructuren met LISREL gesteld kunnen worden. Enkele conclusies uit deze Monte Carlo studie worden aan de hand van een vijftal aspecten toegelicht: convergentie van meest aannemelijke schattingsmethoden, negatieve schattingen van varianties, het effect van het analyseren van correlatiestructuren, het effect van kleine steekproeven, en het effect van scheve verdelingen der geobserveerde variabelen. De praktische consequenties van schendingen van statistische modelassumpties worden beklemtoond.

1. De robuustheidsvraag

De term robuustheid werd als statistisch begrip pas in 1953 door Box geïntroduceerd. Hij noemt een statistische procedure robuust "if its performance is relatively insensitive to departures from the underlying assumptions used to derive that procedure " (p. 318). Wetenschappelijke onderzoekers hebben zich echter met afwijkingen van modelvooronderstellingen bezig gehouden vanaf het eerste moment dat er goed gedefinieerde statistische procedures werden toegepast (vergelijk Stigler, 1973, voor enkele historische notities). In praktische bewoordingen komt de definitie van Box er op neer dat een statistische procedure robuust wordt genoemd tegen gespecificeerde afwijkingen van bepaalde assumpties, wanneer bij het optreden van die afwijkingen in de empirie de conclusies, die op basis van waarnemingen door de onderzoeker worden getrokken, niet wezenlijk anders zouden zijn wanneer wel aan alle assumpties zou zijn voldaan.

Bij het gebruik van statistische procedures, of bij het opstellen van een statistisch of kanstheoretisch model, zijn er altijd bepaalde vooronderstellingen in het geding. Vaak betreft dat kenmerken van de steekproefelementen en de verdelingseigenschappen der waargenomen variabelen. Bij de analyse van empirische gegevens zou de keuze van welk statistisch model dan ook vrijwel altijd de robuustheidsvraag

---

Lezing gehouden op de Statistische Dag, 19 april 1984, Technische Hogeschool te Delft.

A. Boomsma, Vakgroep Statistiek & Meettheorie, Rijksuniversiteit Groningen, Oude Boteringestraat 23, 9712 GC Groningen.

moeten oproepen. En deze vraag luidt allereerst: in welke mate zijn de waargenomen gegevens in overeenstemming met de assumpties van het statistisch model dat zal worden gehanteerd? En indien het antwoord op deze vraag negatief is, daarop aansluitend een tweede vraag, namelijk: hoe erg is het wanneer niet in een geconstateerde mate aan specifieke assumpties is voldaan? Anders gezegd: in welke richting en in welke mate zijn de geconstateerde afwijkingen der assumpties van invloed op de conclusies, die na de statistische analyse van de steekproef-gegevens worden getrokken?

Zeker is dat bij het hanteren van statistische methoden de robuustheidsvraag van centraal belang is; zie Cox & Hinkley, (1974, p. 81) voor een formelere benadering van de probleemstelling. En dat belang heeft niet alleen theoretisch, maar bovenal praktisch gewicht. Gebruikers van statistische procedures zouden moeten weten, en willen vaak ook weten, wat bij benadering het effect van zekere assumptieschendingen op hun statistische conclusies is. Soms evenwel kunnen ze dat volstrekt niet weten, omdat er te weinig of niets bekend is over het effect van specifieke schendingen der vooronderstellingen. In zo'n situatie is het praktisch nuttig en theoretisch interessant om doelgericht robuustheidsonderzoek te gaan uitvoeren. Dit is vooral nuttig onder twee voorwaarden: ten eerste wanneer het een statistische procedure betreft die vaak wordt toegepast, en ten tweede wanneer zich daarbij vaak duidelijke, dat wil zeggen bij herhaling grote, assumptieschendingen voordoen. Onder zulke omstandigheden is het praktische belang van gericht robuustheidsonderzoek voldoende gewaarborgd.

Wat betreft de analyse van covariantiestructuren met behulp van LISREL (Jöreskog & Sörbom, 1981) is aan de twee hierboven genoemde voorwaarden ruimschoots voldaan. Het statistisch model LISREL wordt veelvuldig gebruikt (zie bijvoorbeeld The Journal of Marketing Research (1982; Volume 19) met het speciale november-nummer over de analyse van causale, structurele modellen). Bovendien treden in die toepassingen herhaaldelijk assumptieschendingen op, terwijl over de effecten daarvan op de statistische conclusies tot voor kort zeer weinig bekend was.

In de volgende paragraaf wordt kort ingegaan op de meest essentiële vooronderstellingen bij het gebruik van LISREL. Vervolgens wordt de opzet van een robuustheidsonderzoek met empirische methoden besproken, terwijl tenslotte een aantal praktische aanbevelingen naar aanleiding van zo'n onderzoek aan de hand van vijf aspecten wordt toegelicht. Voor een volledig overzicht van de verrichte robuustheidsstudie wordt naar Boomsma (1983) verwezen. Dit artikel heeft primair tot doel het belang en een mogelijke opzet van robuustheidsresearch aan te geven en het praktisch nut ervan aan de hand van enkele resultaten te illustreren.

2. De statistische assumpties van LISREL

Alvorens te specificeren waartegen de robuustheid van LISREL zou moeten worden onderzocht, is het wenselijk de essentie van het

statistisch model beknopt weer te geven.

Onderzoekers die het LISREL model hanteren zijn in principe geïnteresseerd in theoretische relaties tussen verschijnselen; zoals bekend is de term LISREL een afkorting voor LIneaire Structurele RELaties. Het opstellen van een theoretisch model voor de soms complexe onderlinge samenhangen tussen een geheel van verschijnselen vormt het uitgangspunt van dit type research, dat algemeen bekend staat als analyse van covariantiestructuren. Vervolgens wordt het gespecificeerde theoretische model middels een steekproef van waarnemingen aan de werkelijkheid getoetst. Het statistisch model LISREL, zoals dat door Karl Jöreskog (1970, 1973) werd ontwikkeld, is voor het bestuderen van covariantiestructuren een krachtig en veel gebruikt analytisch hulpmiddel. En niet in de laatste plaats omdat het de onderzoeker in staat stelt modelmatig rekening te houden met meet-onnauwkeurigheden van de geobserveerde variabelen.

In een analyse van covariantiestructuren op basis van een LISREL model staan een drietal statistische aspecten centraal.

(i) schattingen van parameters

De mate waarin er een, vaak gerichte, samenhang tussen verschijnselen bestaat, of de mate van het effect dat het ene verschijnsel op een ander verschijnsel heeft in het theoretische model, kan uitgedrukt worden in een getal. De statistische analyse van het model is er in de eerste plaats op gericht om een schatting te krijgen van dat soort getallen, die abstract modelparameters worden genoemd en hierna worden aangeduid met ω_i. Schattingen van deze parameters worden aangegeven als $\hat{\omega}_i$.

Afhankelijk onder meer van de statistische vooronderstellingen die een onderzoeker bereid is te aanvaarden, kan voor de wijze van schatten gebruik worden gemaakt van een meest aannemelijke schattingsmethode, van een kleinste kwadraten schattingsprocedure, of van de methode der zo genoemde instrumentele variabelen. Hier wordt een beperking gemaakt tot meest aannemelijke schattingen.

(ii) schattingen van standaardfouten

Bij gegeven schattingen van de onbekende modelparameters is het met de meest aannemelijke schattingsprocedure mogelijk een benadering te geven van de betrouwbaarheid van die schattingen. De standaardfout van een parameterschatting, aangeduid als $se_{\hat{\omega}_i}$, kan worden geschat.

(iii) schattingen van de mate van aanpassing

Bovenal is het van belang te beschikken over indicaties voor de discrepantie tussen het theoretische model, dat als uitgangspunt van de analyse werd gekozen, en de empirische bevindingen zoals waargenomen in een steekproef. Zij moeten een indruk geven hoe aannemelijk de geobserveerde steekproefcovarianties $\underset{\sim}{S}$ zijn, gegeven de schatting van de covarianties in de populatie, $\underset{\sim}{\Sigma}$, die het geschatte theoretische model representeren.

Bij de meest aannemelijke schattingsprocedure wordt als uitgangspunt voor die discrepantie (of voor de mate van aanpassing) vaak een schatting van de aannemelijkheidsratio genomen. Een functie van die schatting, -2 log aannemelijkheidsratio, staat bekend als de chi-kwadraat toetsingsgrootheid voor aanpassing (in het Engels, chi-

square test for goodness of fit).

De vooronderstellingen bij het gebruik van de meest aannemelijke schattingsprocedure voor het algemene LISREL model kunnen thans beknopt worden weergegeven.

Uitgegaan wordt van een steekproef van N elementen bij elk waarvan k stochastische variabelen zijn gemeten. Dit resulteert in een gegevensmatrix $\underset{\sim}{Z}$ (N * k). Teneinde statistische uitspraken te doen over de populatieparameters en over de adequaatheid van het gekozen theoretische populatiemodel zijn drie assumpties cruciaal.

1. Aangenomen wordt dat de rijen van $\underset{\sim}{Z}$ onafhankelijk verdeeld zijn. Er is sprake van een aselecte steekproef uit een gespecificeerde populatie waarover de onderzoeker uitspraken wil doen.
2. Er wordt aangenomen dat iedere rij van $\underset{\sim}{Z}$ een multivariate normale verdeling heeft met dezelfde vector van k gemiddelden $\underset{\sim}{\mu}$ en dezelfde covariantiematrix $\underset{\sim}{\Sigma}$(k * k).
3. Er wordt vanuit gegaan dat de steekproefomvang N groot is. De meest aannemelijke schattingstheorie is hier asymptotische theorie.

Wanneer aan de assumpties van onafhankelijkheid en van multivariate normaliteit wordt voldaan, is de steekproevenverdeling van de geschatte modelparameters ω_i, alsmede de steekproevenverdeling van de chi-kwadraat maat voor aanpassing, in grote steekproeven bij benadering bekend. Hoe zien deze steekproevenverdelingen er uit? Bij grote steekproeven heeft de gestandaardiseerde parameterschatting $\omega_i^* = (\hat{\omega}_i - \omega_i)/se\hat{\omega}_i$ bij benadering een standaard normale verdeling. Eveneens bij grote steekproeven heeft de schatting voor de mate van aanpassing χ^2, die is gebaseerd op de aannemelijkheidsratio, bij benadering een chi-kwadraat verdeling met $k(k + 1)/2 - s$ vrijheidsgraden, waarbij s het aantal onafhankelijke modelparameters is dat geschat moet worden.

3. Twee robuustheidsvragen bij LISREL

Gelet op de assumpties die in de vorige paragraaf werden genoemd zijn er drie algemene robuustheidsvragen te stellen, namelijk die naar het effect van schendingen tegen respectievelijk onafhankelijkheid, multivariate normaliteit en grote steekproeven. In het onderzoek van Boomsma (1983) werd ervan afgezien het effect van afhankelijke steekproeftrekkingen te onderzoeken. Vanuit theoretisch oogpunt leek het met name van belang te weten hoe de schattingen van parameters met hun bijbehorende schattingen van standaardfouten en de schattingen van de chi-kwadraat maat voor aanpassing zich gedragen, wanneer niet aan de assumptie van grote steekproeven en die van multivariate normaliteit is voldaan. Iets concreter luidt dan de vraagstelling, hoe de steekproevenverdelingen van deze schattingen er uit zien als een van beide assumpties in een te specificeren mate geschonden is.

Het praktisch belang van deze probleemstelling is, zoals eerder aangeduid, evident als wordt bedacht dat het bij de toepassingen van LISREL veelvuldig voorkomt dat de steekproefgrootte klein is, terwijl de geobserveerde variabelen ook meermalen duidelijke niet-normale

verdelingseigenschappen bezitten. Boomsma (1983, Chapter 5) geeft een overzicht van enkele empirische verdelingen uit concreet, toegepast sociaal-wetenschappelijk onderzoek.

In het zojuist genoemde proefschrift is, kortweg, getracht een antwoord te geven op de volgende twee vragen: a) hoe robuust is LISREL tegen het gebruik van kleine steekproeven? b) hoe robuust is LISREL tegen het gebruik van niet-normaal verdeelde variabelen?

4. Onderzoeksopzet: een Monte Carlo studie

Om beide robuustheidsvragen te beantwoorden zijn Monte Carlo procedures gehanteerd (een analytische benadering van de onderzoeksdoelstelling werd in eerste instantie te lastig gevonden). Met behulp van deze empirische methoden werden er, op basis van een steekproefomvang N, herhaalde malen pseudo-aselecte steekproef covariantiematrices $\tilde{S}$ gegenereerd uit een gespecificeerde populatie, die vervolgens gebruikt zijn als steekproefinvoer voor LISREL analyses.

Met betrekking tot de eerste robuustheidsvraag a) zijn deze covariantiematrices gebaseerd op kleine steekproeven uit een multivariate normale populatieverdeling met covariantiestructuur $\tilde{\Sigma}$. Voor de tweede robuustheidsvraag b) zijn de steekproeven $\tilde{S}$ gebaseerd op trekkingen uit een multivariate niet-normale populatieverdeling (discrete verdelingen met een bepaalde mate van scheefheid, en een steekproefomvang van N = 400). Gegeven deze steekproefomvang van 400, met zoals in paragraaf 5 zal blijken robuuste eigenschappen, komt de onderzoeksopzet praktisch neer op een gescheiden aanpak voor beide robuustheidsvragen, zonder interactie-effecten.

De nadere precisering van een dergelijk onderzoek vergt een groot aantal beslissingen, waarvan de belangrijkste hieronder kort worden genoemd.

(i) De robuustheid tegen kleine steekproeven werd bestudeerd voor een steekproefomvang van N = 25,50,100,200,400.

(ii) De robuustheid tegen niet-normaliteit werd gespecificeerd door allereerst uit te gaan van discrete stochastische variabelen. Vervolgens werd de mate van niet-normaliteit gedefinieerd door het aantal categorieën van deze discrete, numerieke variabelen, en door een bepaalde mate van scheefheid van hun verdelingen. Concreet komt het er op neer dat het gedrag van LISREL is bekeken wanneer in plaats van normaal verdeelde variabelen, multinomiaal verdeelde variabelen worden gebruikt, waarvan het aantal categorieën (de mate van discreetheid) en de mate van scheefheid betrekkelijk eenvoudig zijn te variëren.

Belangrijk is dat het telkens mogelijk was alle niet-normale variaties in het onderzoeksplan te vergelijken met de corresponderende normale variant (NM), die zonder verdelingsschendingen bij een steekproef-omvang van 400. Bij de evaluatie van zulke vergelijkingen zal worden gesproken van een categoriseringseffect en van een scheefheidseffect (zie paragraaf 5.5).

(iii) Het aantal replicaties NR bedraagt telkens 300. Dat is het

aantal keren dat, bij een gegeven schending van een der assumpties, steekproefcovariantiematrices $\tilde{S}$ met LISREL zijn geanalyseerd. De empirische steekproevenverdelingen van de schattingen, die met de theoretische, asymptotische, verdelingen worden vergeleken, zijn daarmee steeds op 300 waarnemingen gebaseerd.
Op welke wijze de empirische steekproevenverdelingen met hun corresponderende theoretische verdelingen zijn vergeleken wordt door Boomsma (1983, Chapter 3) uitvoerig besproken; ook in dat opzicht werden niet-arbitraire beslissingen genomen.
(iv) Een volgende, met name voor de generaliseerbaarheid van de conclusies, niet te verwaarlozen beslissing in het onderzoeksplan betreft de keuze van de covariantiestructuur modellen waaronder assumptieschendingen worden bestudeerd. Gekozen werd voor enkele empirische modellen, die in literatuur veelvuldig zijn besproken, en voor een aantal theoretische factor-analyse modellen. Voor details verwijzen we naar Boomsma (1983, Chapter 4 en Chapter 7).
(v) Op de rekentechnische aspecten van het genereren van steekproef-covariantiematrices $\tilde{S}$ onder de condities van respectievelijk normaliteit en niet-normaliteit wordt hier niet ingegaan. De steekproeftrekkingen zijn verricht met behulp van IMSL-routines (IMSL, 1982). De trekkingsprocedure is voor beide robuustheidsvragen in principe gelijk; voor de niet-normaliteitsvraag zijn specifieke transformaties van gegenereerde variabelen vereist.
Essentieel in de onderzoeksopzet is wel dat steeds bekend is hoe de covariantiestructuur Σ er in de populatie uitziet. De grootte van de modelparameters in de populatie zijn bekend, evenals de bijbehorende asymptotische schattingen van hun standaardfout bij een gegeven steekproefomvang.
Wat betreft de startwaarden voor de LISREL analyses zijn telkens de bekende waarden van de populatieparameters gekozen.

Tot zover enkele beslissingen met betrekking tot de Monte Carlo onderzoeksopzet. Nu enkele uitkomsten van de studie.

5. Resultaten

In plaats van hier een min of meer volledige samenvatting van het uitgevoerde robuustheidsonderzoek te geven, wordt aan de hand van een vijftal algemene resultaten geïllustreerd dat dit soort Monte Carlo studies voor toepassingen van de onderzochte analysetechniek praktische consequenties kan hebben, met name wat betreft de keuze van de steekproefomvang en de keuze van geobserveerde variabelen.

5.1 Convergentie van schattingsprocedures

Om de meest aannemelijke schattingen van de onbekende model-parameters te vinden maakt LISREL gebruik van iteratieve schattings-procedures. In LISREL V (Jöreskog & Sörbom, 1981) is dat een methode van steilste afdaling in combinatie met het algorithme van Davidon, Fletcher & Powell. Deze procedures kennen convergentie- en stop-

criteria. Een van de problemen waar een gebruiker maar ook een robuustheidsonderzoeker van LISREL mee geconfronteerd kan worden, is dat bij een willekeurige steekproef $\underset{\sim}{S}$ aan het convergentiecriterium, EPS, voor alle parameterschattingen niet binnen 250 iteraties is voldaan (zie Gruvaeus & Jöreskog, 1970, voor details).

In de robuustheidstudie werd besloten, wanneer dit probleem zich zou voordoen de parameteroplossing van de betreffende steekproef als niet-convergerend te beschouwen. Zo'n steekproef van covarianties wordt dan terzijde geschoven, terwijl er vervolgens een nieuwe steekproef wordt geanalyseerd tot er uiteindelijk 300 replicaties zijn gevonden met een convergerende parameteroplossing.

Vanuit een praktisch gezichtspunt is het van belang te weten, hoe vaak en onder welke omstandigheden het convergentieprobleem zich voordoet. In Tabel 1 staat voor vier modellen het percentage niet-convergerende oplossingen bij verschillende steekproefgroottes.

Tabel 1. Het percentage niet-convergerende oplossingen, uitgedrukt als $100X/(300 + X)$, waarbij X het aantal niet-convergerende oplossingen is voordat de 300-ste convergerende is gevonden. Een lege cel betekent 0%.

	steekproefgrootte				
Model	25	50	100	200	400
1	22.1%	5.7%	0.7%		
2	6.5%	0.7%			
3US	46.7%	28.4%	12.8%	2.0%	
3CS	55.0%	35.3%	16.4%	2.9%	

In het algemeen neemt het percentage niet-convergerende oplossingen af als de steekproefomvang groter wordt. Een steekproefgrootte van 400 was voor alle bestudeerde modellen groot genoeg om iedere replicatie binnen 250 iteraties aan het convergentiecriterium te laten voldoen.

Uit Tabel 1 blijkt ook duidelijk dat de mate van convergentie afhankelijk is van het theoretische model. Wat betreft factor-analyse modellen neemt de convergentie af, (i) als de covariantiewaarden in de populatie dichter bij nul liggen, en (ii) als de ratio (aantal geobserveerde variabelen)/(aantal factoren) kleiner wordt.

Praktisch gesproken is de beste preventie ter vermijding van convergentieproblemen het kiezen van een steekproefomvang N die niet klein is. Daarbij lijkt N = 200 voor de meeste modellen afdoende.

Bij toepassingen van LISREL is het mogelijk na 250 iteraties de dan gevonden parameterschattingen te gebruiken als startwaarden voor een verdere analyse, in de hoop dat de volgende 250 iteraties toereikend zijn om het convergentiecriterium te bereiken. Uit enkele uitgevoerde deelstudies bleek echter, dat het soms zelfs onmogelijk is tot een

convergerende oplossing te komen, wanneer er geen maximum aan het aantal toegestane iteraties wordt gesteld (zie Boomsma, 1984, voor details).

5.2 Negatieve schattingen van varianties

Vaak moeten er bij de analyse van covariantiestructuren schattingen van varianties worden gemaakt. Een mogelijk probleem dat dan optreedt, is dat bij een gegeven steekproef de LISREL oplossing resulteert in één of meer negatieve schattingen van varianties. In de Engelstalige literatuur staat dit probleem bekend als dat van een "improper solution", terwijl het specifiek in de factor-analyse ook wel een Heywood-geval wordt genoemd.

De huidige versies van LISREL leggen tijdens het iteratiesproces niet automatisch zodanige beperkingen aan het model op, dat de schattingen van varianties onder de parameters niet negatief kunnen worden. Lee (1980) laat zien dat het in principe mogelijk is van LISREL een minimaliseringsprogramma te maken waarbij restricties voor bepaalde schattingen zijn aan te brengen. Recente publicaties op dit terrein zijn van Kelderman (1985) en van Rindskopf (1984).

Ook voor dit probleem van oneigenlijke oplossingen is het vooral van belang te weten hoe vaak en onder welke condities het zich voordoet, omdat dat tot praktische aanbevelingen kan voeren. In Tabel 2 wordt voor enkele modellen bij vijf steekproefgroottes een overzicht gegeven van het percentage replicaties dat tot negatieve schattingen van varianties leidt.

Tabel 2. Het percentage replicaties (NR = 300) waarbij negatieve schattingen van varianties optreden. Een lege cel betekent 0%.

	steekproefgrootte				
Model	25	50	100	200	400
1	64.0%	40.0%	19.7%	4.3%	
2	46.3%	8.7%	0.3%		
3US	51.0%	41.3%	22.0%	10.7%	2.7%
4UM	22.3%	5.0%			
4UL	7.3%	0.3%			

Uit deze tabel blijkt dat het percentage negatieve variantieschattingen afneemt als de steekproefomvang groter wordt, terwijl de grootte ervan sterk modelafhankelijk is. Daarnaast hangt de frequentie waarin negatieve schattingen voorkomen af van de grootte van de populatiewaarden dezer varianties. Bij onderling vergelijkbare modellen en ook binnen een specifiek model met meerdere varianties, is

de frequentie groter bij parameters met een relatief kleine populatiewaarde (zie Boomsma, 1984, voor details).

In concrete toepassingen is het altijd mogelijk varianties, die tot een negatieve schatting leiden, in het theoretische model op een kleine positieve waarde te fixeren en dan een analyse met het aldus gewijzigde model bij dezelfde steekproef $\underset{\sim}{S}$ uit te voeren. Het is evenwel aantrekkelijker vooraf te kiezen voor een steekproefomvang die met enige waarschijnlijkheid niet tot een oneigenlijke oplossing leidt. Een grootte van N = 50 lijkt daarvoor een ondergrens.

5.3 Het analyseren van correlatiestructuren

Met LISREL is het in principe mogelijk correlatiematrices in plaats van covariantiematrices te analyseren. Het gebruik van die optie betekent bij de meeste modellen dat op de diagonaal van de geschatte populatiecovariantiematrix enen staan: diag $\underset{\sim}{\Sigma} = \underset{\sim}{I}$. Omdat er in de statistische theorie van wordt uitgegaan dat er een steekproef $\underset{\sim}{S}$ afkomstig uit een Wishart verdeling wordt geanalyseerd, is de analyse van correlatiematrices, hoewel dat per definitie gestandaardiseerde covariantiematrices zijn, niet zonder consequenties.

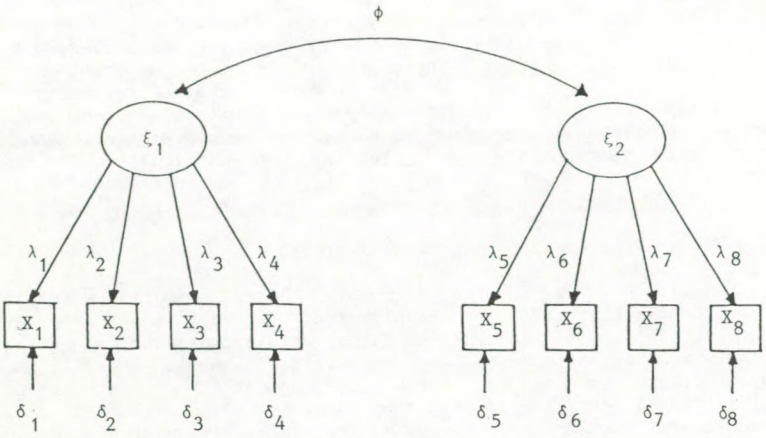
De analyse van correlatiematrices heeft vooral effect op de schattingen van de covarianties tussen parameterschattingen, dat wil zeggen op de geschatte standaardfouten en de schattingen van de product-moment correlaties tussen parameterschattingen. Voor veel modelparameters geeft LISREL een overschatting van de standaardfouten als er correlatiematrices worden geanalyseerd; dat wil zeggen dat de standaardafwijkingen van de geobserveerde parameterschattingen duidelijk kleiner zijn dan de verwachte standaardfouten.

Dit algemene resultaat wordt geïllustreerd aan de hand van een factor-analytisch model, dat gecodeerd is als Model 4CM. Een schematische presentatie daarvan wordt in Figuur 1 gegeven. Er zijn twee gecorreleerde factoren ξ , elk met vier geobserveerde variabelen X_i, terwijl er sprake is van een zogenaamde eenvoudige structuur. De factorladingen λ_i hebben de volgende populatiewaarden: $\lambda_1 = \lambda_2 = \lambda_5 = \lambda_6 = 0.6$, en $\lambda_3 = \lambda_4 = \lambda_7 = \lambda_8 = 0.8$. De varianties van de meetfouten δ_i zijn voor i = 1,...,8 gedefinieerd als $\theta_i = 1 - \lambda_i^2$. De correlatie ϕ tussen de factoren is in de populatie gelijk aan 0.3.

Uit de vergelijking van de analyse van covariantiestructuren ($\underset{\sim}{S}$) met die van correlatiestructuren ($\underset{\sim}{R}$) blijkt dan in Tabel 3, dat de empirische standaardafwijkingen van de gestandaardiseerde parameterschattingen ω_i^* bij analyse van $\underset{\sim}{S}$ vrij dicht in de buurt van de verwachte waarde van 1 liggen, terwijl ze bij analyse van $\underset{\sim}{R}$ duidelijk kleiner zijn dan 1. Zoals uit de tabel valt af te lezen hangt dit resultaat nauwelijks af van de steekproefomvang; voor N = 1000 kunnen soortgelijke uitkomsten worden verwacht. Bij andere modellen zijn vergelijkbare resultaten gevonden, hoewel het afwijkend effect niet voor alle parameters tot een overschatting van de standaardfout leidt.

Wanneer de empirische standaardafwijking wordt vergeleken met de asymptotisch geschatte standaardfout in de populatie, kan het verschil tussen analyses van $\underset{\sim}{S}$ en $\underset{\sim}{R}$ nog dramatischer worden geïllustreerd (zie bijvoorbeeld Boomsma, 1983, Table 4.4.7).

Een aantoonbaar effect van de hierboven genoemde bevindingen is



Figuur 1. Model 4CM: een factor-analytisch model met twee gecorreleerde factoren.

Tabel 3. 100 keer de standaardafwijking van gestandaardiseerde parameterschattingen $\omega_i^* = (\hat{\omega}_i - \omega_i)/se_{\hat{\omega}_i}$, minus 1. Model 4CM. Een tabelwaarde van 0 betekent dat er een perfecte overeenstemming is tussen de empirische resultaten (NR = 300) en de theoretische verwachtingen. $\underset{\sim}{S}$ betekent analyse van covariantiematrices, $\underset{\sim}{R}$ analyse van corresponderende correlatiematrices. N is de steekproefgrootte.

	N = 100		N = 200		N = 400	
ω_i	$\underset{\sim}{S}$	$\underset{\sim}{R}$	$\underset{\sim}{S}$	$\underset{\sim}{R}$	$\underset{\sim}{S}$	$\underset{\sim}{R}$
λ_1	5	-19	6	-20	1	-21
λ_2	0	-23	4	-22	1	-21
λ_3	0	-38	10	-35	8	-35
λ_4	3	-35	-1	-37	1	-43
λ_5	5	-21	0	-24	-5	-25
λ_6	-2	-25	0	-23	-5	-29
λ_7	-3	-40	3	-35	5	-40
λ_8	1	-38	0	-36	0	-40

onder meer dat de geschatte betrouwbaarheidsintervallen voor populatieparameters niet in overeenstemming zijn met de theoretische verwachtingen, ongeacht de steekproefgrootte. Opgemerkt zij dat onder bepaalde condities standaardisatie van $\underline{S}$ geen effect heeft op de chi-kwadraat maat voor aanpassing.

Voor theoretische besprekingen van het in deze paragraaf behandelde onderwerp verwijzen we naar Lawley & Maxwell (1971) en naar Browne (1982).

Een praktisch advies op basis van de gevonden resultaten is dat wanneer correlatiestructuren worden geanalyseerd de geschatte covarianties tussen parameterschattingen onder alle omstandigheden gewantrouwd moeten worden, omdat er dan voor die schattingen met LISREL niet de nodige correcties worden uitgevoerd.

Samenvattend wordt geconcludeerd dat tegen het automatisme om met LISREL correlatiestructuren te analyseren ernstig moet worden gewaarschuwd.

5.4 Het effect van kleine steekproeven

In deze paragraaf worden een viertal aspecten van de robuustheid van LISREL tegen kleine steekproeven bij normale verdelingen aangestipt, te weten zuiverheid, variantie, betrouwbaarheidsintervallen en mate van aanpassing. In deze en in de volgende paragraaf gelden alle resultaten voor analyses van covariantiematrices.

Bij steekproeven van omvang $N > 100$ treden er bij alle onderzochte modellen weinig afwijkingen op met betrekking tot de zuiverheid van de schattingen van parameters en hun bijbehorende standaardfouten; voor $N > 200$ geldt hetzelfde voor de variantie van deze schattingen, afgezien van de effecten van uitbijters en oneigenlijke schattingen.

Wat betreft 95% betrouwbaarheidsintervallen voor populatieparameters ω_i zijn de resultaten bij een steekproefomvang van $N > 100$ voor de meeste modellen volgens verwachting.

Voor de chi-kwadraat toetsingsgrootheid voor aanpassing worden er bij een steekproefomvang van $N < 100$, doorgaans lichte afwijkingen van de theorie gevonden. In het algemeen valt echter te verwachten dat deze aanpassingsmaat redelijk robuust is wanneer $N > 100$. In Tabel 4 wordt dit aan een model voor structurele vergelijkingen (voorbeeld 7 uit Jöreskog & Sörbom, 1981) geïllustreerd.

Uit die tabel blijkt dat voor $N < 100$ de rechterstaart van de empirische steekproevenverdeling duidelijk te dik is: voor $N = 25$ zijn er 17% in plaats van een verwachte 5% chi-kwadraat waarden die groter zijn dan 27.6. Voor de praktijk betekent dit dat een onderzoeker het populatiemodel te vaak ten onrechte zal verwerpen. Bij $N \geq 200$ zijn de afwijkingen voor de maat van aanpassing duidelijk kleiner.

Samenvattend zijn de conclusies met betrekking tot het effect van kleine steekproeven, dat LISREL robuust is als $N \geq 200$ en zeker niet robuust als $N < 100$. Op basis van deze globale grenzen is het bij exploratief modelonderzoek niet aan te raden steekproeven van een kleinere omvang dan 200 te nemen, en bepaald niet wanneer de onderzoeker bij het voorhanden zijn van één enkele steekproef waarde hecht aan enige vorm van kruisvalidatie.

Tabel 4. Kenmerken van de empirische steekproevenverdeling van de chi-kwadraat toetsingsgrootheid voor aanpassing in vergelijking met de verwachte waarden onder een chi-kwadraat verdeling met 17 vrijheidsgraden.
Model 2, NR = 300; gemid. = gemiddelde, std.afw. = standaardafwijking, scheefh. = scheefheid, $\chi^2_{17} > 27.6$ = het percentage chi-kwadraat waarden groter dan het 95% quantiel.

geobserveerde minus verwachte waarde

N	gemid.	std.afw.	scheefh.	$\chi^2_{17} > 27.6$
25	4.5	1.6	0.0	12%
50	2.2	0.9	0.3	5%
100	1.2	0.2	0.1	3%
200	1.0	0.6	0.3	2%
400	0.6	0.1	0.2	1%

	gemid.	std.afw.	scheefh.	$\chi^2_{17} > 27.6$
verwachte waarde	17.0	5.8	0.7	5%
std. fout van geobserveerde waarde; NR = 300	0.4	0.3	0.2	1%

5.5 Het effect van niet-normaliteit

Dezelfde vier statistische aspecten die in de vorige paragraaf zijn besproken worden nu op het effect van niet-normaliteit bekeken. De algemene resultaten worden toegelicht aan de hand van het eerder genoemde Model 4CM (zie Figuur 1).

Allereerst wordt in Tabel 5 een overzicht gegeven van de bij dat model gehanteerde variaties in mate van niet-normaliteit. Het aantal categorieën C van de 8 variabelen varieert van 2 to 5. Dit aantal is telkens gelijk voor verschillende variaties in de mate van scheefheid S. Onder S0 zijn de univariate verdelingen van de discrete variabelen symmetrisch; onder S3 is er sprake van relatief extreme scheefheid.

Het effect van categorisering wordt onderzocht door de resultaten die verkregen zijn onder de corresponderende normale (NM) variant (geen normaliteitsschendingen, N = 400) te vergelijken met de

Tabel 5. Het aantal categorieën C en variaties in de mate van scheefheid S van de discrete geobserveerde variabelen X_i bij de onderzoeksopzet van Model 4CM.

variabele	C	mate van scheefheid S0	S1	S2	S3
X_1	5	0	0.375	0.75	1.5
X_2	2	0	0.750	1.50	3.0
X_3	3	0	0.500	1.00	2.0
X_4	2	0	0.250	0.50	1.0
X_5	5	0	1.000	2.00	4.0
X_6	2	0	0.375	0.75	1.5
X_7	3	0	1.000	2.00	4.0
X_8	2	0	0.750	1.50	3.0

resultaten onder S0, bij gegeven discreetheid C. Het effect van scheefheid wordt bekeken door NM te vergelijken met resultaten verkregen onder een toenemende mate van scheefheid, van S0 tot en met S3.

In het algemeen werden er geen afwijkingen in de zuiverheid van schattingen van parameters en hun bijbehorende standaardfouten gevonden, maar wel, en soms grote afwijkingen in de varianties van de parameterschattingen. De empirische varianties van de parameterschattingen zijn vaak ofwel te klein (bij kleine scheefheden), ofwel te groot (bij matig tot grote scheefheden). Er is hier sprake van een categoriserings- en een scheefheidseffect. Dit wordt geïllustreerd in Tabel 6 aan de hand van enkele parameters uit Model 4CM (zie Figuur 1).

De afwijkingen met betrekking tot de varianties van de parameterschattingen hebben een direct effect op de betrouwbaarheidsintervallen voor modelparameters. Uit Tabel 7 blijkt dat bij extreme scheefheden (S3) de populatiewaarden van de parameters veel te weinig door de 95% intervallen worden bedekt. Het scheefheidseffect is hier zeer pregnant.

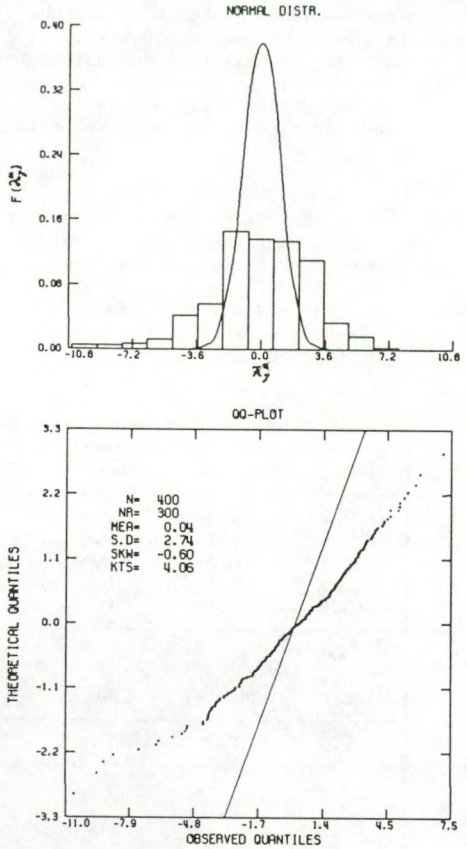
Door een grafische weergave van de empirische steekproevenverdeling van de parameters in vergelijking met de standaard normale verdeling is het effect van scheefheid op een directe manier te visualiseren. Uit Figuur 2 blijkt hoe sterk de geobserveerde steekproevenverdeling van een gestandaardiseerde parameter, in dit geval factorlading λ^*_7, bij extreme scheefheden (S3) kan afwijken van zijn theoretische verdeling [zie Gnanadesikan, 1977, over QQ- (quantiel-quantiel-) plots].

Tabel 6. 100 keer de standaardafwijking van gestandaardiseerde parameterschattingen $\omega_{ij}^{*} = (\hat{\omega}_{ij} - \omega_i)/se_{\hat{\omega}_{ij}}$, minus 1.
Model 4CM, NR = 300, N = 400. Een tabelwaarde van 0 betekent dat er een perfecte overeenstemming is tussen de empirische resultaten en de theoretische verwachtingen.

		mate van scheefheid			
ω_i	NM	S0	S1	S2	S3
λ_1	1	-4	0	-1	32
λ_2	6	-16	-8	11	73
λ_3	6	-7	-1	8	50
λ_5	2	-1	-18	53	152
θ_1	5	-3	4	3	36
θ_2	2	-3	2	10	59
θ_3	-2	2	9	17	53
θ_5	1	-4	11	63	195

Tabel 7. Het percentage 95% betrouwbaarheidsintervallen voor ω_i, $\hat{\omega}_{ij} \pm 1.96\, se_{\hat{\omega}_{ij}}$, minus 5, dat de populatiewaarde ω_i niet bedekt; $j = 1,\ldots,300$.
Model 4CM, NR = 300, N = 400. Toelichting: het getal -2 betekent dat 3% van de 300 intervallen de populatiewaarde ω_i niet bedekt.

		mate van scheefheid			
ω_i	NM	S0	S1	S2	S3
λ_1	0	-2	-2	0	9
λ_2	2	-4	-1	2	20
λ_3	1	-1	1	3	14
λ_5	0	1	5	16	35
θ_1	1	-1	1	0	10
θ_2	1	-1	2	4	17
θ_3	1	-1	2	5	15
θ_5	0	-2	3	18	45



Figuur 2. Histogram (met standaard normale dichtheidsfunctie) en bijbehorende QQ-plot van de gestandaardiseerde parameter λ_7^*. Model 4CM. Mate van scheefheid S3.

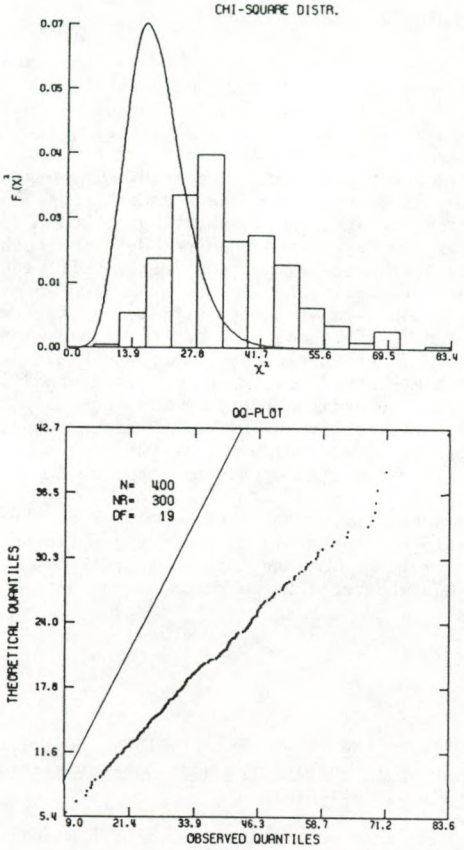
Over het effect van niet-normaliteit op de chi-kwadraat grootheid voor aanpassing kan worden opgemerkt dat daarbij vrijwel geen categoriseringseffect optreedt, maar wel een scheefheidseffect. Tabel 8 laat zien dat naarmate de scheefheid van de variabelen toeneemt de empirische steekproevenverdeling meer naar rechts schuift. Bij Model 4CM is de mate van absolute scheefheid onder S3 gemiddeld 2.5. Bij zulke schendingen van normaliteit heeft dat als effect dat in plaats van de theoretisch verwachte 5% niet minder dan 54% van de geobserveerde chi-kwadraat waarden groter zijn dan het 95% quantiel van 30.1. Voor praktische toepassingen betekent dit dat onder dergelijke condities van niet-normaliteit de nulhypothese van de modeltoets ten onrechte te vaak wordt verworpen. De onderzoeker zal bij deze scheefheden der variabelen moeilijk een juiste beslissing kunnen nemen omtrent de aannemelijkheid van het theoretische model.

Tabel 8. Kenmerken van de empirische steekproevenverdeling van de chi-kwadraat toetsingsgrootheid voor aanpassing in vergelijking met de verwachte waarden onder een chi-kwadraat verdeling met 19 vrijheidsgraden.
Model 4CM, NR = 300, N = 400; gemid. = gemiddelde, std.afw. = standaardafwijking, scheefh. = scheefheid, $\chi^2_{19} > 30.1$ = het percentage chi-kwadraat waarden groter dan het 95% quantiel.

mate van normaliteit	geobserveerde minus verwachte waarde: gemid.	std.afw.	scheefh.	$\chi^2_{19} > 30.1$
NM	0.4	0.5	0.3	0%
S0	0.6	0.5	-0.0	3%
S1	1.3	0.5	0.4	4%
S2	4.9	1.4	0.1	4%
S3	15.4	6.1	0.1	49%

	gemid.	std.afw.	scheefh.	$\chi^2_{19} > 30.1$
verwachte waarde	19.0	6.2	0.6	5%
std. fout van geobserveerde waarde; NR = 300	0.4	0.3	0.2	1%

Om het zojuist besprokene grafisch te illustreren wordt in Figuur 3 de geobserveerde steekproevenverdeling van de aanpassingsmaat uitgezet tegen de theoretisch verwachte verdelingsvorm bij relatief scheve geobserveerde variabelen.



Figuur 3. Histogram (met de X^2_{19} dichtheidsfunctie) en bijbehorende QQ-plot van de chi-kwadraat maat voor aanpassing. Model 4CM. Mate van scheefheid S3.

Op basis van alle bestudeerde modellen kunnen de conclusies met betrekking tot het effect van niet-normaliteit als volgt worden samengevat: LISREL is niet robuust tegen variabelen met relatief scheve verdelingen, maar vrij robuust tegen het categoriseren van symmetrisch verdeelde variabelen.

6. Nadere (robuustheids)vragen

Er zijn veel aspecten van dergelijk robuustheidsonderzoek die nadere bespreking behoeven. In dit korte bestek is het echter niet mogelijk daarop in te gaan; Boomsma (1983, Chapter 8) geeft een aanzet tot discussie. Voor een kritische reflectie over de gevonden resultaten of ter stimulering van voortgezet onderzoek kunnen bijvoorbeeld de volgende vragen worden gesteld. In welke mate zijn de gevonden resultaten te generaliseren naar andere modellen dan de hier onderzochte? Zouden er met een schattingsmethode van kleinste kwadraten bij kleine steekproeven ongeveer dezelfde resultaten zijn gevonden? Zijn de uitkomsten voor niet-normaliteit generaliseerbaar naar continue scheve verdelingen? Is het praktisch aanvaardbaar om scheve variabelen tot symmetrische variabelen te transformeren, teneinde een scheefheidseffect mogelijk te omzeilen? Zijn er robuuste schattingsprocedures denkbaar die het effect van kleine steekproeven of zelfs dat van niet-normaliteit kunnen verminderen? Onder welke omstandigheden is het verdedigbaar om, zoals met LISREL mogelijk is, polychore en polyseriële correlaties te gebruiken? En tenslotte een tot nu toe onbesproken robuustheidsvraag: wat is het effect van het foutief specificeren van het theoretische model op de conclusies van de onderzoeker?

Zonder deze vragen hier trachten te beantwoorden zou er voor de statistische praktijk althans iets gewonnen zijn wanneer zich bij onderzoekers, ook met betrekking tot het gebruik van LISREL, een continu besef van kwetsbaarheid tegen schendingen van modelassumpties zou ontwikkelen.

Referenties

Boomsma, A. (1983) On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality. Amsterdam: Sociometric Research Foundation.

Boomsma, A. (1984) Facts and failures of LISREL modeling. (accepted by Psychometrika)

Browne, M.W. (1982) Covariance structures. In D.M. Hawkins (Ed.) Topics in applied multivariate analysis. Cambridge: Cambridge University Press, 72-141.

Cox, D.R., & Hinkley, D.V. (1974) Theoretical statistics. London: Chapman & Hall.

Gnanadesikan, R. (1977) Methods for statistical data analysis of multivariate observations. New York: Wiley.

IMSL (1982) IMSL Library. Reference Manual. (4 vols.; 9th ed.). Houston, Texas: International Mathematical and Statistical Libraries.

Jöreskog, K.G. (1970) A general method for analysis of covariance structures. Biometrika, 57, 239-251.

Jöreskog, K.G. (1973) A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press, 85-112.

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.

Kelderman, H. (1985) Lisrel models for inequality constraints in factor and regression models. In P.F. Cuttance & J.R. Ecob (Eds.) Structural modeling. Cambridge: Cambridge University Press. (forthcoming)

Lawley, D.N., & Maxwell, A.E. (1971) Factor analysis as a statistical method. (2nd ed.). London: Butterworth.

Lee, S.Y. (1980) Estimation of covariance structure models with parameters subject to functional constraints. Psychometrika, 45, 309-324.

Rindskopf, D. (1984) Using phantom and imaginary latent variables to parameterize constraints in linear structural models. Psychometrika, 49, 37-47.

Stigler, S.M. (1973) Simon Newcomb, Percy Daniell, and the history of robust estimation, 1885-1920. Journal of the American Statistical Association, 68, 872-879.