

IDENTIFIKATIE IN LISREL

E.S. TAN

SAMENVATTING

Het interpreteren van LISREL-modellen dient te gebeuren nadat volledig inzicht is verschaft omtrent de identificeerbaarheid van deze modellen. Een procedure is voorgesteld die de gebruiker/ster in staat stelt een gespecificeerd model te identificeren. Deze procedure gaat uit van een model dat reeds door de gebruiker/ster gesplitst is in een structureel- en een meetgedeelte. De identifikatiestatus van elk gedeelte apart wordt vastgesteld. De verschillende kondities die de gebruiker/ster dient te controleren, geeft hem/haar een beter inzicht in de mogelijkheden een LISREL-model te identificeren.

correspondentieadres
E.S. TAN
Rijksuniversiteit Limburg
Medische Informatica en Statistiek
Postbus 616
6200 MD Maastricht
tel.: 043-434410 tst. 46

Inleiding

Voordat een gegeven LISREL model aangepast wordt aan beschikbare data, is het noodzakelijk de identificeerbaarheid van het gespecificeerd model te onderzoeken. Daar het LISREL-VI programma, speciaal hiervoor ontwikkeld door K.G. Jöreskog en D. Sörbom (10), in staat is parameters te schatten van modellen die niet geïdentificeerd zijn, dient het interpreteren van resultaten slechts te gebeuren nadat een volledig inzicht is verschaft over de identificeerbaarheid van het model of van een deelcollectie van de parameters. (zie hiervoor Jöreskog (9)).

Dit artikel is bedoeld voor de reguliere LISREL-gebruiker, die in de schattings- en interpretatie fase zeker wil zijn dat het door hem gespecificeerd model ook geïdentificeerd is. De hier aangeboden condities geven de gebruiker in veel gevallen inzicht in de extra benodigde informatie die voldoende is een (niet geïdentificeerd) model te identificeren; informatie welke in veel gevallen niet noodzakelijk is, maar relatief eenvoudig in het model in te bouwen is. Bijvoorbeeld door toevoeging van extra exogenen en/of endogenen, of door plausibele veronderstellingen te doen omtrent de relatie tussen enkele variabelen en/of storingstermen. Op praktische gronden zijn uit veelal bestaande literatuur enkele relatief eenvoudig te hanteren condities gegeven die voldoende zijn voor identificatie van speciale, in de praktijk vaak voorkomende, klassen van modellen. Het is niet de bedoeling een volledig beeld te geven van wat er in de literatuur op dit gebied reeds bekend is. Het artikel biedt de gebruiker wel een instrument aan dat hem in bepaalde gevallen de identificatie status oplevert, en in andere gevallen inzicht verschaft in de mogelijkheden het model te identificeren.

LISREL SPECIFIKATIE.

LISREL is het meest algemene model geschikt als beschrijving van relaties in een gegeven causaal systeem.

Het LISREL model is gesplitst in twee gedeelten.

1. Het structureel gedeelte.

Hier worden de gegeven causale relaties tussen de theoretische variabelen gespecificeerd.

2. Het meetmodel gedeelte.

Hier worden de relaties tussen de theoretische en de waargenomen variabelen gespecificeerd.

ad 1. Het structureel gedeelte

$$\bar{\eta} = B\bar{\eta} + \Gamma\bar{\xi} + \bar{\zeta}.$$

$\bar{\eta}$ een $p \times 1$ vector van endogene variabelen.

B een $p \times p$ matrix van structurele parameters.

Γ een $p \times q$ matrix van structurele parameters.

$\bar{\xi}$ een $q \times 1$ vector van exogene variabelen.

$\bar{\zeta}$ een $p \times 1$ vector van storingstermen.

$$E\bar{\xi} = \bar{0}, E\bar{\zeta} = 0, E\bar{\eta} = \bar{0}$$

ad 2. Het meetmodel gedeelte.

$$\bar{X} = \Lambda_x \bar{\xi} + \bar{\delta}$$

$$\bar{Y} = \Lambda_y \bar{\eta} + \bar{e}$$

$\bar{X}$ een $m \times 1$ en $\bar{Y}$ een $n \times 1$ vector van waargenomen variabelen.

Λ_x en Λ_y een $m \times q$ resp. $n \times p$ matrices van coëfficiënten.

$\bar{\delta}$ en $\bar{e}$ een $m \times 1$ resp. $n \times 1$ vector van meetfouten.

$$E\bar{\delta} = \bar{0}, E\bar{e} = \bar{0}, E\bar{e} = \bar{0}.$$

In LISREL zijn $\bar{X}$ en $\bar{Y}$ beschouwd als een realisatie uit een $(m+n)$ -variate, parametrische verdeling met covariantie matrix Σ . We nemen aan dat $\bar{\xi}$, $\bar{\delta}$, $\bar{\eta}$, en $\bar{e}$ normaal verdeeld zijn met verwachtingswaarde nul, zodat $\bar{X}$ en $\bar{Y}$ normaal verdeeld zijn en uitgedrukt als deviaties van hun verwachtingswaarde.

LISREL veronderstelt:

1. ξ ongecorreleerd aan $\bar{\xi}$.
2. $\bar{e}$ ongecorreleerd aan $\bar{\eta}$.
3. $\bar{\delta}$ ongecorreleerd aan $\bar{\xi}$.
4. ξ , $\bar{e}$ en $\bar{\delta}$ zijn onderling ongecorreleerd.
5. B heeft nullen op de diagonaal en is niet singulier.

(1) - (4) zijn gebruikelijke aannames. De nullen op de diagonaal van B zijn normalisaties.

Er zijn twee situaties denkbaar, die leiden tot een singuliere matrix B.

De eerste, minst interessante, situatie doet zich voor wanneer het gespecificeerd model niet logisch consistent is, dat wil zeggen dat de structurele vergelijkingen strijdig zijn. Dit probleem kan eenvoudig opgelost worden door de vergelijkingen te controleren op strijdigheid.

De tweede mogelijke oorzaak van een singuliere B is een overbodige vergelijking. In dit geval is er minstens 1 vergelijking te schrijven als een lineaire combinatie van de overige vergelijkingen. In de praktijk komt dit neer op het herformuleren van de causale processen die aan de modelspecificatie ten grondslag liggen.

DEFINITIE IDENTIFIKATIE.

We veronderstellen dat de simultane verdeling van $\bar{X}$ en $\bar{Y}$ volledig bepaald is door de parameters van het gespecificeerd model, zeg $\bar{\theta}$.

Een definitie die geschikt is voor toepassingen met behulp van LISREL-VI achtige programma's, maakt gebruik van de (populatie) covariantie-matrix Σ .

Σ is als functie van $\bar{\theta}$ op unieke wijze bepaald. Het gespecificeerd model heet nu geïdentificeerd indien geen tweetal $\bar{\theta}_1$ en $\bar{\theta}_2$, $\bar{\theta}_1 \neq \bar{\theta}_2$ dezelfde Σ genereren.

LISREL veronderstelt dat $(\bar{X}, \bar{Y})$ multivariaat normaal verdeeld is. Deze verdeling is op een lokatie na volledig gekarakteriseerd wordt door Σ .

Voordat de parameters van een gegeven LISREL model geschat worden (met behulp van beschikbare schattingsprocedures), moet de identificeerbaarheid van het model reeds onderzocht zijn.

Interpretatie van niet geïdentificeerde modellen dient met de nodige voorzichtigheid te gebeuren, daar enkele van de parameters niet op unieke wijze vastliggen. (Jöreskog (9)).

Identifikatie van een LISREL-model betekent in feite identificatie van alle parameters van dat model. Enkele parameters van een niet geïdentificeerd model kunnen echter wel uniek bepaald zijn.

Als de interesse uitgaat naar een specifieke deelcollectie van de parameters, bijvoorbeeld een deelcollectie bestaande uit de elementen van B en/of Γ , dan is het niet nodig het volledig model geïdentificeerd te krijgen. Programma's zoals LISREL-VI kunnen niet aangeven welke parameters niet geïdentificeerd zijn of welke stappen ondernomen moeten worden om een model geïdentificeerd te maken. Inzicht over de identificeerbaarheid van een gegeven LISREL-model en in het bijzonder de identificatie van de parameters van zo'n model, is daarom onontbeerlijk.

Identifikatie hangt af van de keuze van een model. Om het identifikatie-probleem te onderzoeken van een gegeven model moeten de covarianties als functies van de model-parameters opgelost worden. Deze vergelijkingen worden covariantie-vergelijkingen genoemd. De informatie die nodig is om het stelsel vergelijkingen op te lossen, wordt gehaald uit de covariantie matrix Σ . De covariantie - vergelijkingen zijn in het algemeen niet - lineaire functies van de model-parameters, zodat oplossingen in de meeste gevallen (voorzover deze bestaan) moeilijk te bepalen zijn (Jöreskog (10)).

Een eenvoudig voorbeeld laat zien op welke manier ongedetermineerdheid optreedt in de parameters.

Beschouw het volgende model

$$y_1 = \gamma_{11} x_1 + \zeta_1$$

voorbeeld 1

$$y_2 = \gamma_{21} x_1 + \gamma_{22} x_2 + \beta_{21} y_1 + \zeta_2$$

$$\gamma_{12} = \text{covariantie } \zeta_1 \text{ en } \zeta_2$$

In dit voorbeeld blijkt het niet mogelijk te zijn β_{21} van Ψ_{12} te scheiden. Deze ongedetermineerdheid kan belangrijke consequenties hebben voor de interpretatie. Additionele informatie is nodig om β_{21} te identificeren.

In het nu volgende is het identificatie probleem vertaald in termen van oplosbaarheid van een stelsel vergelijkingen. Het volledig LISREL-model wordt gesplitst in een structureel- en een meetgedeelte. Een nodig en voldoende conditie wordt gegeven voor structurele modellen zonder restricties met betrekking tot de relaties tussen de storingstermen (Ψ).

Een uitbreiding van deze condities maakt wel gebruik van restricties met betrekking tot Ψ . Een voldoende voorwaarde voor de identificatie van meetmodellen is afkomstig van Howe (7). Deze conditie is met name van toepassing voor modellen met ongecorrleerde errors (uniciteiten) die aan bepaalde voorwaarden voldoen.

Tot slot is een voldoende conditie gegeven voor meetmodellen met gecorrleerde errors. Deze meetmodellen worden door LISREL-gebruikers in longitudinale studies veel toegepast. Het schema aan het eind vormt een samenvatting van de verschillende condities die hier gepresenteerd zijn. De gebruiker zou volgens dit schema de identificatie status van zijn model kunnen vaststellen. Het is echter niet de bedoeling hierdoor de lezer een waterdichte procedure aan te bieden. Het door het schema aangeven stap 2 vergt soms de nodige vindingrijkheid van de gebruiker het model adequaat te splitsen in een structureel en een meetgedeelte. Boven-dien zijn er genoeg voorbeelden van geïdentificeerde modellen die niet voldoen aan de condities die hier gegeven zijn. In die situaties is het raadzaam om per model de identificatie status vast te stellen door de covariantie-vergelijkingen op te lossen. het schema eindigt overigens in dat geval met het teken "S".

1. OPLOSBAARHEID VAN STELSEL COVARIANTIE - VERGELIJKINGEN.

Bij een gegeven LISREL model, bijvoorbeeld het model in voorbeeld 1 (hierna model 1 genoemd), is een eenvoudige eerste controle op identifikatie de oplosbaarheid van het stelsel van covariantie - vergelijkingen van de vorm $\Sigma(\theta) = \Sigma$. θ is dus geïdenti-

ficeerd alleen als dit stelsel uniek oplosbaar is voor iedere Σ . Voor de volledigheid geven we hier het covariantievergelijkingen stelsel.

Herschrijft eerst het LISREL model als volgt:

$$A\bar{\eta} = (I-B)\bar{\eta} = \Gamma\bar{\xi} + \bar{\zeta}$$

$$\bar{X} = \Lambda_X \bar{\xi} + \bar{\delta}$$

$$\bar{Y} = \Lambda_Y \bar{\eta} + \bar{e}$$

met de gebruikelijke aanvullende restricties.

De elementen van A , Γ , Φ , Ψ , Θ_e , Θ_δ , Λ_Y en Λ_X vormen de componenten van $\bar{\Theta}$.

Hierbij zijn

$$\Phi = E(\bar{\xi} \bar{\xi}^t), \Psi = E(\bar{\zeta} \bar{\zeta}^t), \Theta_e = E(\bar{e} \bar{e}^t)$$

$$\text{en } \Theta_\delta = E(\bar{\delta} \bar{\delta}^t)$$

Het covariantie-vergelijkingen stelsel ziet er als volgt uit:

$$E \left[\begin{pmatrix} \bar{Y} \\ \bar{X} \end{pmatrix}, (\bar{Y}^t \bar{X}^t) \right] = \Sigma = \Sigma(\bar{\Theta}) = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}$$

met

$$\Sigma_{YY} = \Lambda_Y A^{-1} \Gamma \Phi \Gamma^t A^{-1t} \Lambda_Y^t + \Sigma_{uu}$$

$$\Sigma_{uu} = \Lambda_Y A^{-1} \Psi A^{-1t} \Lambda_Y^t + \Theta_e$$

$$\Sigma_{XX} = \Lambda_X \Phi \Lambda_X^t + \Theta_\delta$$

$$\Sigma_{YX} = \Lambda_Y A^{-1t} \Gamma \Phi \Lambda_X^t$$

$$\Sigma_{XY} = \Sigma_{YX}$$

We beperken ons tot covariantie-vergelijkingen met Σ niet singulier.

Uit de oplosbaarheids criterium is direkt een nodige voorwaarde af te leiden die als eerste check dient voor identifikatie, namelijk

$1/2(m+n)(m+n+1) \geq k$. Hierbij is

het linker lid gelijk aan het aantal covariantie-vergelijkingen en k aan het aantal te bepalen parameters (onbekenden).

Model 1 bevat 7 onbekenden en 10 covariantie-vergelijkingen, zodat aan het oplosbaarheids-criterium voldaan is.

Het is duidelijk dat wanneer niet aan deze voorwaarde voldaan is, er extra restricties op grond van apriori informatie aan het model opgelegd moeten worden. We beperken ons tot die vorm van restricties, in welke de parameters op nul gezet worden. De reden hiervoor zal later duidelijk worden.

2. SPLITSSEN IN EEN STRUCTUREEL-EN MEETGEDEELTE.

Er bestaan geen nodige en voldoende voorwaarden in de literatuur voor identifikatie van een algemeen LISREL-model.

Geraci (5) geeft nodige voorwaarden voor identifikatie van de elementen van θ_δ , A en Γ in het geval dat $\Lambda_y = I_{p \times p}$, $\Lambda_x = I_{q \times q}$ en θ_δ diagonaal zijn. Voor praktische doeleinden zijn de voorwaarden van Geraci echter moeilijk te hanteren. In veel gevallen zijn de grootheden niet alleen onnauwkeurig gemeten; de schaal waarop gemeten wordt is vaak niet hetzelfde. Robinson (11) geeft voldoende voorwaarden voor algemenere modellen, waarbij de exogene variabelen zonder meetfout gemeten zijn.

In dit artikel wordt het LISREL model gesplitst in een structureel- en een meetgedeelte (zie ook Wiley (12)). Uitgaande van deze splitsing wordt het identifikatie-probleem voor elk gedeelte apart onderzocht. Deze procedure levert voldoende voorwaarden voor de identifikatie van speciale, in de praktijk vaak voorkomende, klassen van modellen. Deze klassen omvatten een groot gedeelte van de in de econometrie gebruikte simultane vergelijkingen stelsels zoals MIMIC-modellen en het Klein's model met gelijkheidsrestricties. De procedure is ook geschikt voor de identifikatie van de meeste, in het LISREL-VI manual genoemde, modellen. Een uitzondering hierop vormt de klasse van de zgn. Simplex-modellen. De te volgen procedure is gebaseerd op het feit dat een gegeven LISREL-model geïdentificeerd is zodra zowel het meetgedeelte als het structureel gedeelte geïdentificeerd zijn.

Het structureel gedeelte is geïdentificeerd met betrekking tot de covariantiematrix van de theoretische variabelen.

Uit de identificatie van het meetgedeelte is deze covariantiematrix echter uniek bepaald, zodat het structureel gedeelte ook geïdentificeerd is met betrekking tot Σ .

Uit de identificatie van de afzonderlijke componenten volgt direct de identificatie van de volledige parameter vector $\bar{\theta}$ (vgl. stelling 1.2 Genugten (4)).

Omdat de variabelen in model 1 zonder meetfout gemeten zijn, bestaat het model alleen uit een structureel gedeelte.

Het LISREL-VI programma veronderstelt dat Σ niet singulier is. Singulariteiten in de covariantiematrix ontstaan indien variabelen perfect gerelateerd zijn aan elkaar. In zulke situaties ontbreken dus de storingstermen. Indien de identiteiten gespecificeerd kunnen worden, dan blijkt het mogelijk te zijn het model, met deze identiteiten, aan het LISREL-VI programma aan te bieden. Identiteiten komen nogal eens voor in econometrische modellen.

Heel nadrukkelijk moet worden gesteld dat de procedure in het algemeen kan leiden tot verschillende voldoende voorwaarden, afhankelijk van de keuze hoe het model in een structureel en een meetgedeelte gesplitst wordt.

A. IDENTIFIKATIE STRUCTUREEL MODEL.

De LISREL-specificatie voor het structureel gedeelte ziet er als volgt uit:

$$A\bar{Y} = \Gamma\bar{X} + \bar{\zeta} \quad \text{met}$$

- A niet singulier
- Ψ symmetrisch niet singulier
- $E(\bar{x}\bar{\zeta}^t) = 0$
- Σ niet singulier.

In de econometrie ook bekend als het "pure-shock contemporaneous" model (Geraci (4)). Een veel voorkomende deelklasse van modellen is de zogenaamde klassen van recursieve modellen.

Een recursief model is een structureel model, waarvoor de vergelijkingen zo gerangschikt kunnen worden opdat A een onder-driehoeksmatrix is en Ψ een diagonaal matrix.

We zullen (nodig en) voldoende voorwaarden geven voor achtereenvolgens

- recursieve modellen
- modellen zonder restricties met betrekking tot Ψ
- modellen met restricties met betrekking tot Ψ

1. RECURSIEVE MODELLEN.

- A is een niet singuliere onderdriehoeksmatrix met enen op de diagonaal vanwege normalisatie restricties.
- Ψ is een niet singuliere diagonaal matrix.

Recursieve modellen zijn speciale gevallen van modellen met restricties ten aanzien van Ψ . Ze hebben de plezierige eigenschap altijd geïdentificeerd te zijn.

Model 1 is geen recursief model omdat Ψ niet diagonaal is. Dat model in matrix vorm ziet er als volgt uit.

$$\begin{pmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

met $\Psi_{12} = \text{cov}(\zeta_1, \zeta_2)$.

Als aangenomen kan worden dat de covariantie tussen y_1 en y_2 volledig verklaard wordt door de relatie met x_1 en x_2 , dan is $\Psi_{12} = 0$, zodat het model geïdentificeerd is.

2. RANGCONDITIE.

Voor modellen met willekeurige A (en normalisatie restricties) en willekeurige niet singuliere Ψ , is bekend (Fisher (3), van der Genugten (4) en Geraci (5)) dat de i-de vergelijking geïdentificeerd is indien de rang van de matrix C_i , bestaande uit de kolommen van (A; Γ) waarvoor de i-de rij 0 is, gelijk is aan n-1; het totaal aantal endogenen minus 1.

De volgende identifikatie regel is hier dus van toepassing.

- (A) rang (C_1) = n-1 is een nodig en voldoende voorwaarde voor identifikatie van de i-de vergelijking van het structureel model zonder restricties ten aanzien van Ψ .

Een eenvoudig na te gaan nodige voorwaarde voor identifikatie van de i-de vergelijking, volgt uit de rangconditie.

- (A 1) Het aantal uitgesloten exogenen in vergelijking i moet minstens gelijk zijn aan het aantal endogenen in deze vergelijking minus 1.

of met betrekking tot het volledig model.

- (A 2) Het totaal aantal uitgesloten endogenen en exogenen moet minstens gelijk aan het totaal aantal endogenen minus 1.

We kunnen bovengenoemde regels op model 1 toepassen.

$$A = \begin{pmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

$$\Psi_{12} = \text{cov}(\zeta_1, \zeta_2)$$

Nu is $C_1 = \begin{pmatrix} 0 & 0 \\ 1 & \gamma_{22} \end{pmatrix}$, zodat de eerste vergelijking geïdentificeerd is.

De tweede vergelijking voldoet echter niet aan voorwaarde (A 1), zodat deze vergelijking niet geïdentificeerd is. Aanvullende informatie is nu nodig. Door de restrictie $\gamma_{21} = 0$ bijvoorbeeld, is identifikatie gewaarborgd mits $\gamma_{11} \neq 0$.

Samenvattend kan m.b.t. het identifikatie probleem van model 1 gesteld worden dat

- zonder enige aanvullende restrictie β_{21} niet uniek bepaald is.
- als $\Psi_{12} = 0$, dan is model 1 recursief en dus geïdentificeerd.

In het bijzonder is β_{21} uniek bepaald.

- Uit de rangconditie volgt, dat β_{21} ook uniek bepaald is indien $\gamma_{21} = 0$ en $\gamma_{11} \neq 0$.

Situaties kunnen zich voordoen waarin geen enkele restrictie met betrekking tot A opgelegd kan worden. In dat geval is een bruikbare nodige voorwaarde voor identifikatie

(A 3) Aantal exogenen moet minstens gelijk zijn aan het aantal endogenen

voorbeeld 2.

$$A\bar{Y} = \Gamma \bar{X} + \bar{\zeta}, \quad \bar{Y} \text{ symmetrisch.}$$

$$\Gamma = \begin{pmatrix} \gamma_{11} & \theta & \vdots \\ \theta & \cdot & \vdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad \begin{array}{l} \gamma_{ii} \neq 0 \\ i=1, \dots, n \leq m \\ * \text{ een willekeurige onder-} \\ \text{matrix} \end{array}$$

m

Dit model voldoet aan de rangconditie (A) en is dus geïdentificeerd.

(Exogene) variabelen die aan precies één endogene variabele gerelateerd zijn heten instrumentele variabelen.

Merk op dat een geïdentificeerd model niet noodzakelijk geïdentificeerd blijft indien aan het model extra restricties worden opgelegd ($\gamma_{11} \neq 0$ is een noodzakelijke voorwaarde voor identifikatie als $\gamma_{21} = 0$). Vanwege het simultane karakter van de vergelijkingen met betrekking tot identifikatie, moeten de restricties op een speciale manier (namelijk een manier die voldoet aan de rangconditie) verdeeld zijn over de vergelijkingen.

Zoals eerder opgemerkt zijn restricties van de vorm $\theta = 0$ wezenlijk anders dan restricties die parameters op een waarde ongelijk aan nul vastzetten. De nodige voorwaarden, welke afgeleid zijn uit de rangconditie, zullen in het algemeen niet meer gelden. In de praktijk komt dit soort van restricties niet vaak voor.

Bovendien kan identifikatie van modellen met zulke vormen van restricties onderzocht worden door eerst zonder deze restricties de identifikatie te onderzoeken. Is het model eenmaal geïdentificeerd, dan blijft dit zo na het vastzetten van bepaalde parameters op een vaste waarde.

Uiteraard met inachtneming van de relevante rangcondities.

3. UITGEBREIDE RANGCONDITIE.

Indien restricties zijn opgelegd met betrekking tot Ψ , blijft de genoemde rangconditie (A) voldoende voor identifikatie, maar deze is niet nodig. In het bijzonder zijn modellen met singuliere Ψ geïdentificeerd zodra aan de rangconditie (A) is voldaan. De nodige voorwaarden A_1 , A_2 en A_3 zijn echter niet meer geldig. De recursieve modellen lijken volgens rangconditie (A) niet geïdentificeerd. Het recursief model uit het LISREL-VI manual bevat zelfs een vergelijking waarvan alle exogene-en endogene variabelen aanwezig zijn.

We zullen hier een algemene regel geven die voldoende is voor een willekeurig stelsel van structurele vergelijkingen. Algemener, omdat de klasse van modellen die volgens deze regel geïdentificeerd zijn, de recursieve- en alle andere modellen omvat welke volgens de rangconditie (A) geïdentificeerd zijn.

In E.A. Hanushek en J.E. Jackson (6) is een voorbeeld gegeven van een geïdentificeerd model, dankzij restricties ten aanzien van Ψ (regulier).

De hier te geven conditie (B) is in feite een formalisatie van het voorbeeld uit Hanushek en Jackson.

Veronderstel dat $(p-1)$ vergelijkingen (zeg de eerste $(p-1)$, $p > 1$) geïdentificeerd zijn, dan (zie appendix A)

(B) voldoende voor identificatie van de p -de vergelijking is dat de rang van de matrix (Zeg: D_p) bestaande uit de kolommen van A , Γ en de eerste $(p-1)$ kolommen van Ψ waarvoor de p -de rij nul is, gelijk is aan $n-1$.

Een eenvoudig te hanteren voldoende voorwaarde voor de identificatie van een structureel model met willekeurige Γ is een speciaal geval van konditie B nl.

(B1) De p -de vergelijking is geïdentificeerd als het aantal storingstermen uit de eerste $(p-1)$ vergelijkingen die ongecorrigeerd zijn aan de storingsterm van de p -de vergelijking gelijk is aan het aantal endogenen in deze vergelijking minus één.

De identifikatie van recursieve modellen kan ook met behulp van konditie (B1) bevestigd worden. Beschouw namelijk zo'n recursief model.

$$B = \begin{pmatrix} 1 & 0 \\ * & 1 \end{pmatrix}, \quad \Psi = \begin{pmatrix} \Psi_{11} & 0 \\ 0 & \Psi_{nn} \end{pmatrix} \quad \text{niet singulier.}$$

De eerste vergelijking voldoet aan de rangconditie (A) en is dus geïdentificeerd. De tweede vergelijking heeft twee endogenen, zodat de restrictie $\Psi_{21} = 0$ voldoende is voor identificatie. Identifikatie van de derde vergelijking volgt uit de identifikatie van de eerste en de tweede vergelijking en het feit dat $\Psi_{31} = \Psi_{32} = 0$.

Voortzetting van deze procedure leidt uiteindelijk tot de identifikatie van alle vergelijkingen, en dus van het volledig model.

B. IDENTIFIKATIE MEETMODEL.

De LISREL-specificatie van het meetgedeelte ziet er als volgt uit

$$\bar{Y} = \Lambda_Y \bar{\eta} + \bar{\epsilon}$$

$$E(\bar{\eta} \bar{\epsilon}^t) = 0, \quad \Sigma_{YY} \quad \text{niet singulier.}$$

De regels die hierna genoemd worden kunnen toegepast worden op de afzonderlijke meetmodellen (m.b.t. $\bar{\eta}$ resp $\bar{\xi}$) of op de gecombineerde meetgedeelte.

$$\bar{Y} = \Lambda_Y \bar{\eta} + \bar{\epsilon} \quad \text{en} \quad \bar{X} = \Lambda_X \bar{\xi} + \bar{\delta} \quad \text{kunnen namelijk gecombineerd}$$

worden tot $\bar{Z} = \Lambda \bar{\Theta} + \bar{\epsilon}$

$$\text{met} \quad \bar{Z} = \begin{pmatrix} \bar{Y} \\ \bar{X} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_Y & 0 \\ 0 & \Lambda_X \end{pmatrix} \quad \text{en} \quad \bar{\epsilon} = \begin{pmatrix} \bar{\epsilon} \\ \bar{\delta} \end{pmatrix}.$$

In het algemeen leiden de afzonderlijke en gecombineerde aanpak tot verschillende resultaten. We zullen hier echter niet verder op ingaan, maar ons beperken tot het geven van voldoende kondi-

ties voor meetmodellen van de vorm $\bar{Y} = \Lambda_Y \bar{\eta} + \bar{e}$. (Beide aanpak leiden tot dezelfde vorm.)

Het meetmodel is in factor analytische context reeds door verscheidene mensen onderzocht. Anderson en Rubin (1) gaven enkele nodige voorwaarden voor identifikatie. De meeste daarvan zijn alleen theoretisch van belang. Bovendien eist een factor analytisch model alleen identifikatie op een rotatie-factor na. De elementen van Λ_{yy} behoeven niet noodzakelijk uniek bepaald.

In LISREL zijn juist deze elementen belangrijk omdat ze een maat geven voor de betrouwbaarheid van de Y-variabelen als meetinstrumenten van de bijbehorende theoretische variabelen.

Het covariantie-vergelijkingen model is

$$\Sigma_{yy} = \Lambda_Y \Sigma_{\eta\eta} \Lambda_Y^t + \Theta_e \quad (II)$$

$\Sigma_{\eta\eta}$ een symmetrische matrix met

$$\text{diag}(\Sigma_{\eta\eta}) = I_{p \times p}$$

Deze restrictie is nodig om een eenheidsmaat voor de theoretische variabelen vast te zetten. Dit kan ook bereikt worden door elke variabele te schalen volgens één van de y-variabelen door de desbetreffende λ 's op één gelijk te stellen.

1. MODELLEN DIE VOLDOEN AAN DE VOLDOENDE VOORWAARDE VAN ANDERSON EN RUBIN

In eerste instantie zullen we Θ_e diagonaal veronderstellen. Een nuttige voldoende voorwaarde die een unieke Θ_e garandeert is afkomstig van Anderson en Rubin (1) (stelling 5.1).

De voorwaarde luidt als volgt:

Θ_e (varianties van de meetfouten) is uniek bepaald, als er van Λ_Y twee deel-matrices overblijven van rang p, na weglating van een willekeurige rij uit deze matrix.

voorbeeld 3

$n=7, p=3$

Meetmodel met

$$\Lambda_Y = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & \lambda_4 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_5 & \lambda_6 \\ 0 & 0 & \lambda_7 \\ 0 & 0 & 1 \end{pmatrix}$$

θ_e diagonaal

bijvoorbeeld het weglaten van de laatste rij levert

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda_7 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & \lambda_4 & 0 \\ 0 & \lambda_5 & \lambda_6 \end{pmatrix} = \begin{pmatrix} \lambda_Y^1 \\ | \\ \lambda_Y^2 \end{pmatrix} \quad \text{met}$$

$\text{rang}(\Lambda_Y^1) = \text{rang}(\Lambda_Y^2) = 3$, mits $\lambda_7 \neq 0$, $\lambda_2 \neq 0$, $\lambda_4 \neq 0$ en $\lambda_6 \neq 0$; met soortgelijke resultaten bij weglating van de andere rijen.

In het algemeen is het een tijdrovende zaak de bovengenoemde voorwaarde van Anderson en Rubin te controleren. Meestal is aan de voorwaarde voldaan indien het aantal waargenomen variabelen veel groter is dan het aantal theoretische, niet waargenomen, variabelen in het model; in ieder geval moet gelden.

$2p+1 \leq n$, p aantal theoretische variabelen en
 n aantal waargenomen variabelen

De elementen van Λ hoeven echter niet uniek bepaald te zijn. Een voldoende voorwaarde voor uniciteit van Λ -parameters die voor LISREL-toepassingen erg nuttig blijkt te zijn is die van Howe (7).

- (c) Voldoende voor identifikatie van een type 1 model is rang $\Lambda_s = p - 1$. Hierbij is Λ_s de submatrix bestaande uit de rijen van Λ waarvoor de s -de kolom 0 is.

Het model van voorbeeld 6 voldoet aan de voorwaarde van Howe en is dus geïdentificeerd. Als $\lambda_6 = 0$, dan geldt de voorwaarde van Anderson en Rubin niet meer. De voorwaarde van Howe is nu niet meer relevant. In Bollen en Jöreskog (2) is een voorbeeld gegeven van een niet geïdentificeerd meetmodel dat wel voldoet aan de voorwaarde van Howe.

2. MODELLEN MET GECORRELEERDE ERRORS

Een klasse van meetmodellen die vaak in LISREL wordt toegepast is de volgende

$$\begin{matrix} (I) \\ \bar{Y} = D|\bar{\eta} + \bar{e}, \bar{e} = (\bar{e}_1, \bar{e}_2, \bar{e}_3)^t \\ \Lambda \quad \quad \bar{e}_1 = (e_{11}, \dots, e_{p1})^t. \end{matrix}$$

I een $p \times p$ identiteit-matrix

D een $p \times p$ diagonaal-matrix

Λ een $r \times p$ matrix met $r > 0$

$$E(\bar{e}_i \bar{e}_j^t) = \begin{cases} 0 & \text{als } i \neq j \\ \theta_i & \text{anders.} \end{cases}$$

Het is aan te tonen, dat deze klasse van modellen geïdentificeerd is, (zie Appendix B) indien minstens één meetfout in de eerste p vergelijkingen ongecorrleerd is aan alle andere meetfouten.

Deze modellen worden in LISREL-context toegepast in longitudinale studies, zoals de 'multi-wave models' (Jöreskog en Sörbom (8)). Gecorreleerde errors kunnen ontstaan wanneer dezelfde variabelen over tijd herhaald worden. Door herinnerings- en andere hertest effecten kunnen de error termen niet zomaar ongecorrleerd verondersteld worden. Als in de specificatie van het model een achtergrond variabele is opgenomen die constant blijft over tijd, dan is het aannemelijk dat de bijbehorende error ongecorrleerd is aan alle andere error termen. In Jöreskog en Sörbom (8) zijn

enkele voorbeelden genoemd van meetmodellen zoals hierboven beschreven.

VOORBEELD 4.

Als $\lambda_6 = 0$, dan voldoet het model van voorbeeld 3 aan de genoemde konditie voor gecorreleerde errors met

$$\Lambda_Y = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline \lambda_2 & 0 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & 0 & \lambda_7 \\ \hline \lambda_3 & \lambda_4 & 0 \end{pmatrix}$$

Omdat θ_e diagonaal is, zijn alle meetfouten ongecorreleerd aan elkaar. De konditie laat zien dat θ_e niet noodzakelijkerwijs diagonaal moet zijn, zodat ook minder restrictieve modellen gespecificeerd kunnen worden.

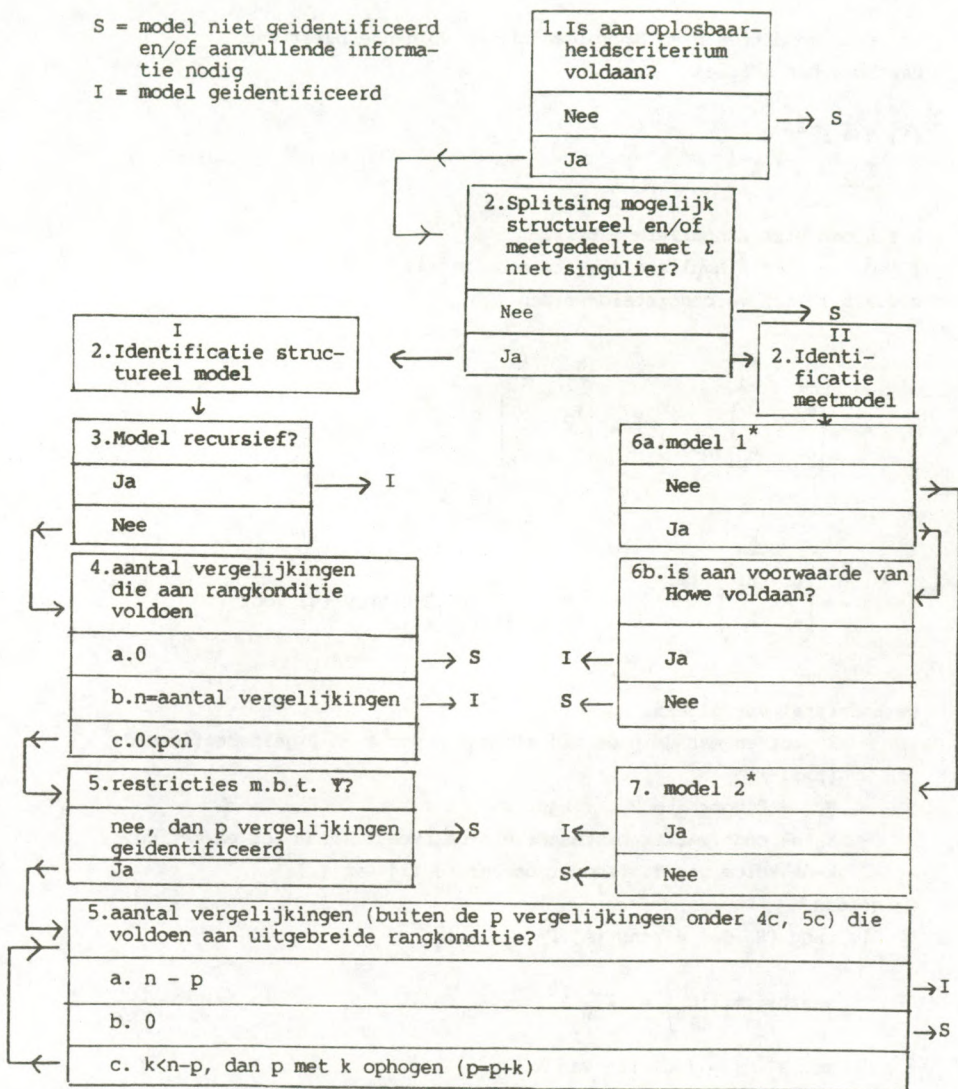
dankbetuiging

De auteur dankt de anonieme referee voor zijn commentaar en suggesties welke tot deze huidige versie leiden.

SCHEMA

S = model niet geïdentificeerd
en/of aanvullende informa-
tie nodig

I = model geïdentificeerd



* model 1 : Modellen die voldoen aan de voorwaarde van Anderson en Rubin. θ_e is dan diagonaal.

model 2 : Modellen^e met gecorreleerde errors.

APPENDIX A

De (rang)kondities B en B1 volgen uit de volgende beweringen.
Beschouw het stelsel

$$(1) \quad \Pi = A^{-1} \Gamma \\ \Gamma_{uu} = A^{-1} \Psi A^{-1} \Gamma$$

A een nxn niet singuliere matrix.

Ψ een nxn niet singuliere symmetrische matrix.

z.d.a.t.s. kan verondersteld worden

$$A = \left(\begin{array}{cc} \overbrace{1}^m & \overbrace{2}^m \\ \alpha_p & 0 \\ * & A_p \end{array} \right)_{n-1}, \quad \Gamma = \left(\begin{array}{cc} \overbrace{1}^k & \overbrace{2}^k \\ \gamma_p & 0 \\ * & \Gamma_p \end{array} \right)$$

$$\Pi = \left(\begin{array}{cc} & \overbrace{2}^k \\ \Pi_1 & \Pi_2 \\ * & * \end{array} \right)_{m_1}$$

(zie ook v/d Genugten (4) voor de schrijfwijze).

veronderstel vervolgens:

- 2^e tot en met de p-de rij elementen van A en Γ geïdentificeerd ($p > 1$)
- $\Psi_{1j} = 0$ voor $j = 2, \dots, k \leq p$
- Ψ_k de ondermatrix bestaande uit de tweede kolom tot en met de k-de kolom van Ψ . (Zonder de eerste rij van Ψ .)

Er gelden

$$1. \text{ rang } (\Pi_2 \quad Q_k) = \text{rang } (A_p \quad \Gamma_p \quad \Psi_k) - m_2$$

$$\text{waarbij } m_1 \left(\begin{array}{c} k-1 \\ Q_k \end{array} \right) = (\Gamma_{uu} A^2, \dots, \Gamma_{uu} A^k).$$

met A^i : = i-de rij van A.

2. De eerste rij elementen van A en Γ zijn uniek bepaald uit Π en

Γ_{uu} indien

a) $\text{rang } (A_p \quad \Gamma_p \quad \Psi_k) = n-1$

b) indien $k = m_1$

bewijs

1. Uit de schrijfwijze van A, Γ en Π volgt direkt dat

$$A^{-1} \begin{pmatrix} 0 & 0 & 0 \\ \underbrace{A_p}_{m_2} & \underbrace{\Gamma_p}_{k_2} & \underbrace{\Psi_k}_k \end{pmatrix} = m_2 \left\{ \begin{pmatrix} 0 & \Pi_2 & Q_k \\ I_d & * & * \end{pmatrix} \right\}$$

van waaruit het gestelde volgt.

2a. Het stelsel (1) levert een unieke oplossing met betrekking tot de eerste rij elementen van A indien het stelsel

$$\begin{cases} \alpha_p \Pi_2 = 0 \\ A^j \Sigma_{uu} A^{1T} = 0 \quad j = 2, \dots, k \text{ een unieke oplossing heeft.} \end{cases}$$

Omdat $A^j \Sigma_{uu} A^{1T} = A^1 (\Sigma_{uu} A^{jT})$, is het stelsel equivalent

$$\text{aan } \begin{cases} (\alpha_p \ 0) \begin{pmatrix} Q_k \\ * \end{pmatrix} = 0 \\ (\alpha_p \ 0) \begin{pmatrix} \Pi_2 \\ * \end{pmatrix} = 0 \end{cases}$$

$$\text{of } \alpha_p (\Pi_2 \ Q_k) = 0.$$

Dit is een stelsel van $k+k_2-1$ lineaire vergelijkingen met m_1-1 onbekenden, en heeft een unieke oplossing alleen als $\text{rang}(\Pi_2 \ Q_k) = m_1-1$.

uit 1. volgt dus dat α_p uniek bepaald is indien $\text{rang}(A_p \ \Gamma_p \ \Psi_k) = n-1$. Omdat $\Gamma = A\Pi$ en $\Psi = A \Sigma_{uu} A^T$ is het bewijs geleverd.

2b. uit het bewijs van 1 volgt direkt

$\text{rang}(A_p \ \Psi_k) = \text{rang } A_p + \text{rang } Q_k$, uit het bewijs van 2 volgt dus de uniciteit als $k-1 = \text{rang } Q_k = m_1-1$, dus $k=m_1$.

APPENDIX B

Meet modellen met gecorreleerde errors zijn geïdentificeerd indien minstens één meetfout in de eerste p vergelijkingen ongecorrleerd is aan alle andere meetfouten.

bewijs:

schrijf $\bar{y} = (\bar{y}_1^t, \bar{y}_2^t, \bar{y}_3^t)^t$ met $\bar{y}_1$ en $\bar{y}_2$ een p -vector.

$$E(\bar{y}_i \bar{y}_i^t) = \Sigma_{ii}, E(\bar{y}_i \bar{y}_j^t) = \Sigma_{ij}, i, j = 1, 2, 3.$$

veronderstel dat de p -de meetfout in het model ongecorrleerd is aan alle andere meetfouten.

$$\begin{aligned} \text{Er geldt: } \Sigma_{11} &= \Sigma_{\eta\eta} + \theta_e \\ &= \begin{pmatrix} w_{11} + \theta_{11} & & \\ & * & \\ w_{p1} & \dots & w_{p-1,1} & w_{pp} + \theta_{pp} \end{pmatrix} \end{aligned} \quad \dots\dots (1)$$

zodat $w_{p1} \dots w_{p-1,1}$ bepaald zijn.

Met

$$\begin{aligned} D &= \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \quad \text{geldt} \\ \Sigma_{21} &= D \Sigma_{\eta\eta} \\ &= \begin{pmatrix} w_{11}\lambda_1 & \dots & w_{1p}\lambda_1 \\ & & w_{2p}\lambda_2 \\ & & \\ w_{p1}\lambda_p & \dots & w_{pp-1}\lambda_{p-1} & w_{pp}\lambda_p \end{pmatrix} \end{aligned} \quad (2)$$

Uit (1) en de laatste rij van de matrix uit (2) volgt dat λ_p en w_{pp} bepaald zijn. Omdat $w_{ij} = w_{ji}$ volgt direkt de identifikatie van D , θ_e en $\Sigma_{\eta\eta}$. Tot slot is Λ bepaald door $\Lambda = \Sigma_{31} \Sigma_{\eta\eta}^{-1}$.

LITERATUUR.

- [1] Anderson T.W. en (1956) Statistical inference in
Rubin H. factor analysis.
Proc. 3rd Berkeley symp. Math.
Stat. Prob.5, 111-150.
- [2] Bollen K.A. en (1985) Uniqueness does not imply
Jöreskog K.G. identification.
Sociological methods and Research
vol 14, 155 - 165.
- [3] Fisher, F.M. (1966) The identification problem in
econometrics.
Mc Graw - Hill, New York.
- [4] Genugten van der B.B. Identification in statistical
(1977) inference.
Statistica Neerlandica vol. 31,
69-91
- [5] Geraci V.J. Identification of simultaneous
(1976) equation models with measurement
error 163-187
(In Aigner D.J. en Goldberger A.S.
Latent variables in socio-
economic models).
- [6] Hanushek E.A. en Statistical methods for social
Jackson J.E. (1982) scientists Hoofdstuk 9 246-281.
Academic Press.
- [7] Howe H.G. (1955) Some contributions to factor
analysis:
Report no. ORNL - 1919 vak ridge
national laboratory.
Oak Ridge, Tenn.
- [8] Jöreskog K.G. en Statistical models and methods
Sörbom D. (1979) for analysis of longitudinal data
in Advances in factor analysis and
structural equation models Hfd. 5.
- [9] Jöreskog K.G. (1981) Analysis of covariance structures.
Scand. J. Statist 8 65-92.
- [10] Jöreskog K.G. en (1984) LISREL-VI
Sörbom D. Analysis of Linear structural
relationships by the method of
maximum likelihood.
User's Guide.

- [11] Robinson P.M. (1977) Identification, estimation and large sample theory for regressions containing unobservable variables 103-119
(In Aigner D.J. en Goldberger A.S. Latent variables in socio-economic models).
- [12] Wiley D.E. (1973) The identification problem for structural equation models with unmeasured variables 69-83
In Goldberger A.S. and Duncan O.D. Structural equation models in the social sciences.

Ontvangen: 11-02-1987

Geaccepteerd: 08-10-1987