

BIJLAGE 1: DE REKENTOETS

ivo	Optellen			Aftrekken			proportie*
	som-type	toets-opgave	proportie*	nivo	som-type	toets-opgave	
3	e+e>	8+7=	.91	3	te-e<	86-5=	.77
4	t+t=	20+20=	.93	4	te-t<	78-20=	.76
5	te+e=	53+6=	.85	5	te-te<	64-21=	.85
6	te+t<	61+20=	.85	6	te-e=	37-7=	.77
7	te+te<	21+77=	.91	7	te-te=	85-25=	.79
		47+52=	.93			97-7=	.79
8	te+e=	71+9=	.81	8	t-e>	70-4=	.57
		52+8=	.85			50-8=	.57
9	te+te=	25+35=	.89	9	te-e>	44-7=	.57
		68+12=	.89			76-9=	.60
0	te+e>	67+6=	.83	10	t-te>	50-17=	.53
		37+8=	.77			90-81=	.62
1	te+te>	17+56=	.77	11	te-te>	92-14=	.51
		69+23=	.89			81-62=	.42
		44+48=	.85			72-54=	.41

toot

de opgaven in de originele toets werden verticaal aangeboden.

= getal bestaand uit alleen een eenheid

= getal bestaand uit alleen een tiental

= getal bestaand uit een tiental en een eenheid

: = zonder tientaloverschrijding

: = op het tiental uitkomend

: = met tientaloverschrijding

* gemiddelde proportie goed gemaakte opgaven (= moeilijkheidsgraad)

Latente budget analyse en onderzoek naar schoolloopbanen

Peter G.M. van der Heijden¹ en Leo J. Th. van der Kamp²

¹Vakgroep Empirische-Theoretische Sociologie, Rijksuniversiteit Utrecht*

²Vakgroep Methoden en Technieken van Psychologisch Onderzoek, Rijksuniversiteit Leiden

ABSTRACT

Latent budget analysis is a new method for the analysis of contingency tables. The contingency table is transformed into a table with conditional proportions that add up to 1 for each row. Such a row with conditional proportions is called a budget. The model estimates a number of latent budgets that have generated the observed budgets. In this paper the usefulness of latent budget analysis for the analysis of school careers will be discussed. The CBS-SMVO-data will be reanalyzed.

1. INLEIDING

Het onderzoek van schoolloopbanen staat de laatste jaren sterk in de belangstelling, en daarmee ook de methodologie van dit onderzoek. Tal van methoden en technieken zijn voorgesteld voor de analyse van schoolloopbaanvariabelen, en de neerslag hiervan is in de onderzoeksliteratuur terug te vinden. Zo verscheen recent in dit tijdschrift een themanummer over longitudinaal onderzoek van schoolloopbanen, terwijl het nummer over multilevel analyse voor dit type onderzoek eveneens van belang is¹.

Een recent ontwikkelde methode waarmee schoolloopbaanvariabelen ook geanalyseerd kunnen worden is latente budget analyse. Latente budget analyse is een methode om kruistabellen te analyseren. De methode is oorspronkelijk voorgesteld voor de analyse van zgn. tijdbudgetten die verzameld worden wanneer een onderzoeker geïnteresseerd is in de wijze waarin objecten of groepen van objecten (bv. groepen van personen ingedeeld naar een bepaald kenmerk, dieren, regio's, landen) hun tijd besteden. Deze tijdbesteding heeft betrekking op een aantal onderscheiden en elkaar niet overlappende activiteiten. De informatie wordt verzameld in een tweeweg matrix met objecten als rijen en activiteiten als kolommen; elk element van de matrix geeft aan de proportie tijd die het betreffende object aan de corresponderende activiteit besteedt. Elke rij van een dergelijke matrix telt op tot 1.00. De rij-elementen worden ook wel de verklarende variabelen genoemd en de kolom-elementen de responsvariabelen.

In dit artikel wordt latente budget analyse (LBA) geïntroduceerd aan de hand van een eenvoudig voorbeeld. De belangrijkste termen worden gedefinieerd en het achterliggende model wordt uiteengezet. Aangegeven wordt op welke wijze de modelparameters geschat kunnen worden, en de parameters geïnterpreteerd. Vervolgens komt de relatie aan de orde van LBA met andere analyse-methoden voor dit type schoolloopbaanvariabelen. Algemeen kan LBA gezien worden als een vorm van latente structuuranalyse; preciezer geformuleerd: LBA is een verbijzondering van latente structuur analyse in de zin dat de manifeste gegevens worden opgevat als een functie van de achterliggende latente variabelen, waarbij het gaat om categorische variabelen. In de toepassing van deze analyse in onderzoek naar schoolloopbanen zijn (longitudinale) categorische variabelen in het geding.

* Vakgroep Empirisch-Theoretische Sociologie, Postbus 80.140, 3508 TC Utrecht.

dit artikel wordt de relatie van LBA met bepaalde vormen van latente structuur analyse, m.n. entele klasse analyse en loglineaire analyse met latente variabelen aangegeven. Tenslotte wordt de hand van gegevens uit het zgn. CBS-SMVO-bestand een secundaire analyse met LBA gevoerd. De resultaten van deze analyse – uitgevoerd op een 4-weg matrix – zullen worden

sproken.
Tot slot van deze inleiding moet nog het volgende opgemerkt worden m.b.t. longitudinaal derzoek in het algemeen en onderzoek naar schoolloopbanen in het bijzonder. Enerzijds kan onderzoek meer 'theory driven' zijn, anderzijds kan de modellering van de manifeste gegevens meer nadruk krijgen. Het laatste stelt dit artikel ten doel zonder daarbij uit te sluiten dat de alyseresultaten tot enig theoretisch inzicht leiden.

2. LATENTE BUDGET ANALYSE: EEN EENVOUDIG VOORBEELD

bel 1 is afkomstig uit het CBS-SMVO bestand. Hierin zijn gegevens te vinden over kinderen 1977 en 1981. In 1977 zaten deze kinderen in de zesde klas van het lager onderwijs. In 1981 rd gekeken welk niveau deze kinderen behaald hadden na vier jaar voortgezet onderwijs. De or ons gebruikte data zijn te vinden in Meester en de Leeuw (1983).

De matrix heeft als rijen 7 klassen van individuen ingedeeld naar hun scores op een intelli- nite test. De gebruikte intelligentietest was de Toets voor Intellectuele Begaaftheid (TIB), n figur-exclusie test die bestaat uit 33 items. Deze toets schijnt overigens verloren te zijn gaan (zie Meester en de Leeuw (1983)). De TIB scores zijn gecodeerd als 1 voor 1 tot 14 ms goed, 2 voor 15 tot 17 items goed, 3 voor 18 tot 20 items goed, 4 voor 21 tot 23 items goed, 5 voor 24 tot 26 items goed, 6 voor 27 tot 29 items goed, en 7 voor 30 tot 33 items goed. De lommen van tabel 1 hebben betrekking op het behaalde eindniveau, d.w.z. op het diploma dat het voorgezet onderwijs is behaald. De categorieën zijn: voortijdig schoolverlaters (drop ts), LBO, MAVO, HAVO, VWO en (M)BO. Meer over deze en andere variabelen is te vinden Meester en de Leeuw (1983, 1985).

In Tabel 1 staan de conditionele prioriteiten om een bepaald eindniveau te halen gegeven dat en een bepaalde TIB-score heeft. Alle kinderen die een bepaalde IQ score hebben worden als t ware verdeeld over de eindniveaus. Wij noemen zo'n verdeling *een budget*. De term 'bud- it' wordt gesuggereerd door financiële budgetten, waar het totale bedrag in prioriteiten over een mal posten wordt verdeeld, of een tijdbudget, waarbij de totaal beschikbare tijd in priorite- rdeeld wordt over een aantal activiteiten. Voor elk budget tellen deze (conditionele) propo- s op tot 1, en daarom tellen de conditionele prioriteiten in Tabel 1 per rij (d.w.z. per TIB-score)) tot 1.

abel 1. Een 2-weg matrix met rijen TIB-klassen en kolommen EIND-niveau. Gegeven zijn de conditionele prioriteiten om een bepaald eindniveau te behalen gegeven de TIB-score.

B-score	EIND-niveau						Totaal	N
	Drop out	LBO	MAVO	HAVO	VWO	(M)BO		
.157	.447	.158	.061	.021	.157	1.000	805	
.101	.387	.204	.098	.026	.184	1.000	1363	
.078	.313	.198	.147	.065	.199	1.000	2803	
.065	.232	.206	.184	.104	.209	1.000	3839	
.044	.178	.204	.218	.165	.191	1.000	3821	
.036	.106	.177	.254	.249	.179	1.000	2602	
.026	.060	.126	.266	.386	.137	1.000	1003	
.063	.226	.192	.188	.142	.189	1.000	16236	

De bestudering van budgetten lijkt vooral zinvol als de kolomvariabele als afhankelijke variabele en de rijvariabele als onafhankelijke variabele kan worden voorgesteld. In dit geval zijn we immers vaak geïnteresseerd in de vraag wat de kans is op elk van de categorieën van de afhankelijke variabele, gegeven de waarde van de rijvariabele. De elementen van de budgetten, de conditionele prioriteiten, zijn schattingen van deze kansen.

Latente budget analyse dient om de bestudering van de geobserveerde budgetten te verge- makkelijken. De vraag die bij LBA gesteld wordt is: hoe kunnen de *waargenomen* budgetten van de rij-categorieën benaderd worden door een gewogen som van *latente* budgetten die bepaald worden door de kolomcategorieën? Het idee is dat we het totale aantal waargenomen budgetten (1 voor elke rij) vervangen door een kleiner aantal nog onbekende latente budgetten, die als het ware de waargenomen budgetten hebben gegenereerd. Deze latente budgetten zijn dan vaak te interpreteren als 'typische' budgetten. Het budget van elke rij is dan voor een bepaalde hoeveel- heid uit elk van deze typische budgetten opgebouwd.

In deze zin vervult LBA als analysetechniek voor kruistabellen een vergelijkbare rol als factor analyse of principale componenten analyse voor de analyse van kwantitatieve variabelen. Bij factor analyse en principale componenten analyse gaat men immers ook op zoek naar typische variabelen, namelijk de factoren, resp. de principale componenten. Alvorens de vraag van latente budget analyse te beantwoorden en in te gaan op de bepaling van de latente budget- ten, zal eerste de notatie aan de orde komen.

In het algemeen bevat een 2-weg tabel als tabel 1 waargenomen prioriteiten p_{ij}/p_{i+} met rijen i ($i=1, \dots, I$) en kolommen j ($j=1, \dots, J$). De grootheden p_{ij}/p_{i+} vormen per rij een budget, en zijn voorwaardelijke waarschijnlijkheden: in ons voorbeeld van tabel 1, gegeven een bepaalde TIB- klasse, met welke prioriteiten wordt een bepaald eindniveau bereikt?

Het latente budget model beoogt zonder verlies aan relevante informatie elk van de I waarge- nomen budgetten te schrijven als een som van T latente budgetten. Er wordt een onderscheid gemaakt tussen waargenomen budget i met elementen p_{ij}/p_{i+} , het theoretische budget met ele- menten π_{ij}/π_{i+} , en het *latente* budget t met elementen π_{jt}^{BX} . De elementen van elk budget tellen op tot 1.00. Het latente budget model stelt voor de theoretische budgetten

$$\frac{\pi_{ij}}{\pi_{i+}} = \sum_{t=1}^T \frac{\pi_{jt}^{AX} \pi_{jt}^{BX}}{\pi_{jt}^{AX} + \pi_{jt}^{BX}} \quad (1)$$

met restricties $\sum_t \pi_{jt}^{AX} = 1$, $\sum_t \pi_{jt}^{BX} = 1$, en $0 \leq \pi_{jt}^{AX} \leq 1$ en $0 \leq \pi_{jt}^{BX} \leq 1$. De parameters π_{jt}^{AX} laten zien hoe elk theoretisch budget i , met elementen π_{ij}/π_{i+} , opgebouwd is uit latente budget- ten, met elementen π_{jt}^{BX} . Ten gevolge van de restricties zijn de parameters te interpreteren als waarschijnlijkheden. Nemen we aan dat de waarnemingen verzameld zijn onder een product- multinomiale verdeling, dan kan het zgn. EM-algoritme gebruikt worden voor het vinden van maximum likelihood schatters van de modelparameters (zie voor een bespreking van de schattingsprocedure bijvoorbeeld de Leeuw, van der Heijden en Verboon, 1990, en van der Heijden, Mooijaart en de Leeuw, 1990).

Voor de gegevens in Tabel 1 blijkt voor $T=1$ $G^2 = 2330.20$ ($df = 30$), $T=2$ $G^2 = 178.35$ ($df = 20$) en voor $T=3$ $G^2 = 12.12$ ($df = 12$). Het model met $T=3$ latente budgetten blijkt niet signifi- cant te zijn en deze oplossing is dan ook acceptabel. De schattingen van de model-parameters staan vermeld in Tabel 2.

Om een adequate beschrijving te geven van de data uit tabel 1 in termen van het latente budget model blijken drie latente budgetten nodig te zijn. Het model met $T=3$ immers wordt op grond van de $G^2 = 12.12$ met $df = 12$ niet verworpen. De waarden van de latente budgetten van laatstgenoemde model staan in Tabel 2, in de laatste drie kolommen. We bespreken eerst de

Tabel 2. Latente budgetten voor T=1, 2 en 3 voor de data uit Tabel 1.

	T=1			T=2			T=3		
	T=1	T=2	T=3	T=1	T=2	T=3	T=1	T=2	T=3
1. drop out	.063	.104	.007	.000	.000	.173	.041	.007	.162
2. LBO	.226	.396	.000	.109	.565	.000	.927	.073	.395
3. MAVO	.192	.212	.166	.323	.131	.000	.788	.212	.484
4. HAVO	.188	.084	.327	.263	.000	.295	.635	.365	.552
5. VWO	.142	.000	.330	.000	.000	.639	.461	.539	.531
5. (M)BO	.189	.204	.162	.306	.130	.025	.270	.730	.434
Budget grootte	1.000	1.000	1.000	1.000	1.000	1.000	.127	.873	.135
	1.000	.571	.429	.468	.310	.222			

TIB-score	T=1			T=2			T=3		
	T=1	T=2	T=3	T=1	T=2	T=3	T=1	T=2	T=3
1	1.000	.959	.041	.173	.665	.162	.173	.665	.162
2	1.000	.927	.073	.395	.497	.108	.395	.497	.108
3	1.000	.788	.212	.484	.386	.130	.484	.386	.130
4	1.000	.635	.365	.552	.289	.158	.552	.289	.158
5	1.000	.461	.539	.531	.230	.380	.531	.230	.380
6	1.000	.270	.730	.434	.186	.380	.434	.186	.380
7	1.000	.127	.873	.135	.207	.658	.135	.207	.658

Propriëties waarmee latente budgetten worden gebruikt:

latente budget parameter schattingen $\hat{\pi}_{jt}^{BX}$. Het eerste budget is het budget van leerlingen die vaker dan gemiddeld naar de MAVO (.323 versus gemiddeld .192), de HAVO (.263 versus .188) en naar het (M)BO (.306 versus .189) gaan, en heeft men nog een redelijke kans om naar het LBO te gaan (.109), al is deze kans lager dan gemiddeld (.226). Hierbij is het gemiddelde de proporties $P_{+j} = \pi_{jt}^{BX}$ voor het model met T=1, zie de eerste kolom van tabel 2; dit zijn de kansen om na vier jaar een bepaald eindniveau te hebben als er geen informatie over de TIB-score zou zijn. In het tweede budget gaat men veel vaker dan gemiddeld naar het LBO (.565 versus .226 gemiddeld), is men ook vaker dan gemiddeld 'drop out' (.173 versus .063), en heeft men ook nog een redelijk grote kans om naar de MAVO (.131) en naar het (M)BO (.130) te gaan. Voor het derde latente budget geldt dat leerlingen aanzienlijk vaker dan gemiddeld naar de HAVO (.295 versus .188) gaan, en ook vaker dan gemiddeld naar het VWO (.639 versus .142). Samengevat, is het eerste budget voornamelijk het budget van het lagere algemeen onderwijs en de 'drop out', en het tweede budget is het budget van het lagere beroepsonderwijs en de 'drop out', en het derde budget is het budget van het hogere algemene onderwijs. Een globale vergelijking van het latente budget model met T=2 met dat met T=3 laat zien dat het eerste latente budget van het model met T=2 te beschrijven is als 'lagere' onderwijsvormen, en het tweede als 'hogere' onderwijsvormen.

Het latente budget model met meer dan een latent budget kan in deze context geïnterpreteerd worden als een MIMIC model (zie Clogg, 1981; Clogg gebruikte deze interpretatie voor de analyse van sociale mobiliteitstabellen). In Figuur 1 is deze model interpretatie schematisch weergegeven. De interpretatie is als volgt. Er worden T latente klassen verondersteld tussen de verklarende en de te verklaren variabele. Elk kind heeft een bepaalde kans om in deze latente

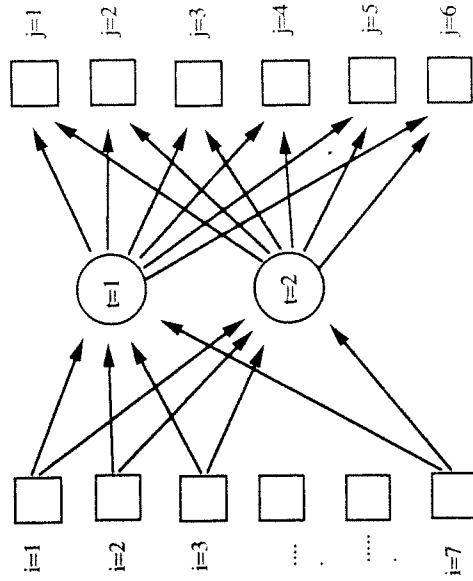


Fig. 1. Latente budgetten analyse als een MIMIC-model: gegeven de toestand waarin kinderen zich bevinden hebben zij een bepaalde kans om naar latente toestand 1 of 2 te gaan; gegeven de toestanden waarin zij terecht zijn gekomen, hebben zij een bepaalde kans om een eindniveau te behalen.

klassen terecht te komen. Deze kansen tellen voor elk kind op tot 1, en zijn bepaald door de categorieën van de verklarende variabele, namelijk zijn/haar TIB-score. Deze kansen zijn gespecificeerd door de π_{jt}^{BX} -parameters. Gegeven dat een kind in een van deze T latente klassen terecht is gekomen, heeft hij/zij J kansen om elk van de J eindniveaus te bereiken. Deze J kansen tellen ook weer op tot 1 en zijn gespecificeerd door de π_{jt}^{BX} -parameters.

Voor andere eigenschappen van LBA verwijzen we naar van der Heijden, Mooijjaart en de Leeuw (1989, 1990), de Leeuw, van der Heijden en Verboon (1990) en de Leeuw en van der Heijden (1991).

3. LATENTE BUDGET ANALYSE EN ANDERE ANALYSEMETHODEN

LBA is nauw gerelateerd aan latente klasse analyse, loglineaire analyse met latente variabelen, en correspondentieanalyse. We bespreken hier de relatie met latente klasse analyse en loglineaire analyse met latente variabelen. LBA is in bepaalde gevallen equivalent aan correspondentieanalyse. We verwijzen voor de relatie met correspondentieanalyse naar van der Heijden, de Leeuw en Mooijjaart (1990) en de Leeuw en van der Heijden (1991).

Latente klasse analyse

Clogg (1981) presenteerde latente budgetanalyse als een herparametrisering van latente klasse analyse. Hierdoor gelden allerlei eigenschappen die gelden voor LBA ook voor latente klasse analyse. Voor een uitgebreide inleiding tot de latente klasse analyse zij verwezen naar McCutcheon (1987) en de daarin aangehaalde literatuur.

Definieer π_{jt}^{BX} als de kans om in latente klasse t te vallen van de latente variabele X, definieer π_{jt}^{AX} als de conditionele kans om in categorie i van variabele A te vallen, gegeven klasse t van

de latente variabele X, en definieer π_{jt}^{BX} als de conditionele kans om in categorie j van variabele B gegeven latente klasse t. Het latente klasse model voor een twee-weg matrix definieert dan een model voor de latente (want niet waargenomen) kansen π_{ij}^{ABX} om in categorie i van variabele A, categorie j van variabele B en categorie t van de latente variabele X te vallen:

$$\pi_{ij}^{ABX} = \pi_{jt}^{AX} \pi_{it}^{BX} \quad (2a)$$

en de latente kansen hebben de volgende relatie met de geobserveerde kansen:

$$\pi_{ij} = \sum_{t=1}^T \pi_{ijt}^{ABX} \quad (2b)$$

Latent budget analyse en latent klasse analyse hebben de parameters π_{jt}^{BX} gemeenschappelijk. De LBA parameters π_{jt}^{AX} kunnen worden afgeleid van de latente klasse analyse parameters π_{jt}^{AX} and π_{jt}^{BX} met behulp van het theorema van Bayes:

$$\pi_{jt}^{AX} = \frac{\pi_{jt}^{AX} \pi_{jt}^{BX}}{\sum_{t=1}^T \pi_{jt}^{AX} \pi_{jt}^{BX}} \quad (3)$$

Latent budget analyse en latente klasse analyse zijn ook vergelijkbaar in termen van de latente kansen π_{ij}^{ABX} . Voor beide modellen geldt voor de kolomparameters dat $\pi_{jt}^{BX} = \pi_{+jt} / \pi_{++t}$; de rij parameters voor latente klasse analyse zijn gerelateerd aan de latente kansen middels $\pi_{jt}^{AX} = \pi_{ABX}^{ABX} / \pi_{+t}^{AX}$ terwijl de rijparameters voor LBA gerelateerd zijn aan de latente kansen π_{jt}^{BX} middels $\pi_{jt}^{AX} = \pi_{ABX}^{ABX} / \pi_{++t}$. In termen van de latente kansen geldt voor beide modellen dat de geobserveerde variabelen A en B conditioneel onafhankelijk zijn gegeven het niveau van de latente variabele X. Bijvoorbeeld, uit (2a) valt af te leiden dat

$$\pi_{ijl}^{ABX} = \pi_{++l}^{ABX} \left(\frac{\pi_{+jl}^{ABX}}{\pi_{++l}^{ABX}} \right) \left(\frac{\pi_{i+1}^{ABX} \pi_{+jl}^{ABX}}{\pi_{++l}^{ABX}} \right) \quad (4)$$

Latente klasse analyse wordt meestal gebruikt in de context van kruistabellen met meer dan twee variabelen. Hier poogt een latente variabele een verklaring te geven voor de samenhang tussen de waargenomen variabelen (zie bijvoorbeeld Goodman, 1974). Er is relatief weinig aandacht geweest voor latente klasse analyse van twee-weg kruistabellen, met als uitzondering de theoretische bijdragen van Good (1969), Gilula (1979, 1983, 1984), Clogg (1981), Goodman (1987), en de Leeuw en van der Heijden (1991), en het werk in sociale mobiliteit en marketing van Marsden (1985), Grover (1987), Grover en Srinivasan (1987) en Luijckx (1987). Voor zover wij weten is de herparameterisering in (1) alleen gebruikt door Clogg (1981) en ons eigen werk (zie de referentielijst).

Een keuze voor ofwel LBA ofwel latente klasse analyse behoort af te hangen van de onderzoeksvraag. Als deze onderzoeksvraag geformuleerd is in termen van een verklarende variabele en een respons variabele, dan lijkt LBA meer geschikt. Als de onderzoeksvraag geformuleerd is in termen van relaties tussen variabelen, dan lijkt latente klasse analyse meer geschikt. In deze zin is het onderscheid tussen LBA en latente klasse analyse vergelijkbaar met het onderscheid tussen het (multinomiale) logit model en het loglineaire model. Wij zullen nog laten zien dat het onderscheid tussen verklarende en responsvariabelen specifieke typen modellen suggereert die niet in de kontekst van latente klasse analyse worden beschouwd.

Het latente budgetten model kan ook worden gepresenteerd als een gerestricteerde versie van het simultane latente klasse analyse model met slechts een manifeste variabele (zie van der Heijden, Mooijaart en de Leeuw, 1989) of als een speciaal geval van het gemengde Markov latente klasse model met een variabele gemeten op een tijdstip (zie van de Pol en Langeheine, 1990). Deze relaties laten zien hoe bestaande software voor simultane latente klasse analyse en het gemengde Markov latente klasse model te gebruiken zijn om schattingen te verkrijgen voor het ongerestricteerde latente klasse model.

Relatie met loglineaire analyse

Formule (4) laat zien dat LBA ook op te vatten is als een loglineair model met een latente variabele:

$$\log \pi_{ijl}^{ABX} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_l^X + \lambda_{il}^{AX} + \lambda_{jl}^{BX} \quad (5)$$

(zie ook Haberman, 1979; Hagenaars, 1990). Dit model toont dat er een directe relatie is tussen A en X, tussen B en X, maar dat A en B niet gerelateerd zijn gegeven X. Deze representatie van LBA is een ingang om verdere restricties op te leggen aan de parameters van het latente budget model. Deze parameters kunnen verder gerestricteerd worden door de parameter λ_{il}^{AX} verder te restricteren. Een voorbeeld hiervan wordt gegeven in sectie 4. Technische details zijn te vinden in van der Heijden, Mooijaart en de Leeuw (1990).

Het voordeel van latente budget analyse

LBA kan gebruikt worden in verschillende situaties. Ten eerste als de onderzoeksvraag daar aanleiding toe geeft, bijvoorbeeld indien men geïnteresseerd is in de conditionele proporties (de geobserveerde budgetten), en men aanneemt dat deze gegeneerd zijn door een kleiner aantal latente budgetten. Deze latente budgetten zijn dan vaak te interpreteren als typische verdelingen. Ten tweede, indien men geïnteresseerd is in de samenhang in de kruistabel, en men inzicht wil verkrijgen in de samenhang tussen de rij- en de kolomcategorieën. Hier zijn natuurlijk meer methoden voor, maar LBA heeft (met latente klasse analyse) het grote voordeel dat de parameters van het model zo eenvoudig te interpreteren zijn. Hierdoor heeft LBA het voordeel dat zij ook tamelijk eenvoudig aan niet-statistisch geschoolden uit te leggen is. Dit is bijvoorbeeld veel moeilijker voor technieken als correspondentie-analyse, loglineaire analyse of hieraan gerelateerde modellen zoals het RC-associatiemodel. Het onderscheid tussen latente budget analyse en latente klasse analyse is hierboven reeds behandeld.

4. SOCIAAL MILIEU AND SECUNDAIR ONDERWIJS

De twee-weg kruistabel die als voorbeeld werd geanalyseerd in sectie 2 was afgeleid uit een vier-weg kruistabel verkregen door op te tellen over de twee van de oorspronkelijke variabelen, namelijk sexe en sociaal milieu. De vier-weg kruistabel is te vinden in de appendix. Het sociaal milieu van de familie is operationeel als het beroep van de vader, en heeft 6 categorieën: 1 is geschoolde en ongeschoolde arbeider, 2 is boer en boerenarbeider, 3 is middenstander, 4 is lagere employee, 5 is middelbare employee, 6 is hogere employee, wetenschappers en vrije beroepen. De variabelen geslacht en sociaal milieu worden door ons opgevat als verklarende variabelen. Daarom worden de categorieën van sexe, sociaal milieu en TIB-score interactief gecodeerd tot 2 x 6 x 7 = 84 rijen, en het bereikte eind-niveau secundair onderwijs levert de 6 kolomcategorieën van de te analyseren matrix. De variabele sexe wordt genoteerd als A met index i, sociaal milieu als C, met index k, TIB als F, met index p, en de respons variabele eindniveau als B, met index j. Het latente budgettenmodel wordt dan

$$\frac{\pi_{ijkp}}{\pi_{ikp}} = \sum_{t=1}^T \frac{ACFX_{tj} BX}{\pi_{ijkp} \pi_{jt}} \quad (6)$$

Tabel 4: Latente budgetten voor T = 2, 3, 4 en 5 voor eindniveau na 4 jaar secundair onderwijs.

	T=2			T=3			T=4		
	i=1	i=2	i=3	i=1	i=2	i=3	i=1	i=2	i=3
1. Drop out	.011	.100	.160	.014	.011	.184	.000	.008	.042
2. LBO	.000	.392	.658	.000	.188	.000	.000	.188	.000
3. MAVO	.171	.208	.121	.090	.325	.124	.198	.059	.357
4. HAVO	.346	.073	.000	.367	.232	.000	.000	.307	.420
5. VWO	.335	.000	.000	.530	.000	.000	.000	.625	.000
6. (M)BO	.137	.228	.061	.000	.432	.000	.614	.000	.181
Budget grootte	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	.423	.577	.343	.267	.389	.266	.225	.227	.283

	T=5					Onafhankelijkheid				
	i=1	i=2	i=3	i=4	i=5	i=T=1				
1. Drop out	.190	.009	.000	.063	.025	.063				
2. LBO	.809	.000	.222	.000	.000	.226				
3. MAVO	.000	.099	.093	.725	.000	.192				
4. HAVO	.000	.223	.000	.000	.878	.188				
5. VWO	.000	.670	.000	.000	.000	.142				
6. (M)BO	.000	.000	.685	.211	.097	.189				
Budget grootte	1.000	1.000	1.000	1.000	1.000	1.000				
	.228	.211	.188	.212	.161					

als we het schrijven als een loglineair model voor de latente kansens π_{ikpj}

$$\log \pi_{ikpj} = \lambda + (\lambda_i^A + \lambda_k^C + \lambda_p^F + \lambda_{ik}^{AC} + \lambda_{ip}^{AF} + \lambda_{kp}^{CF} + \lambda_{ikp}^{ACF}) + \lambda_j^B + \lambda_i^X + \lambda_{ikpt}^{AX} + \lambda_{ikt}^{CX} + \lambda_{ipt}^{FX} + \lambda_{kpt}^{ACX} + \lambda_{kpt}^{CFX} + \lambda_{ikpt}^{ACFX} + \lambda_{jpt}^{BX} \quad (7)$$

et latente budgettenmodel met T=1 (onafhankelijkheid) is equivalent met het loglineaire model voor de geobserveerde variabelen waarbij sexe, sociaal milieu en TIB onderling afhankelijk zijn, maar onafhankelijk van het eindniveau, d.w.z.

$$\log \pi_{ikpj} = \lambda + (\lambda_i^A + \lambda_k^C + \lambda_p^F + \lambda_{ik}^{AC} + \lambda_{ip}^{AF} + \lambda_{kp}^{CF} + \lambda_{ikp}^{ACF}) + \lambda_j^B \quad (8)$$

dit model levert een geschikt ondergrens: in dit model zijn de verklarende variabelen onafhankelijk van de te verklaren variabele. Het model heeft een passing van $G^2 = 4612$ voor 415 vrijheidsgraden.

In eerste instantie bepalen we het aantal latente klassen T dat nodig is om een adequate beschrijving van de data te geven. In Tabel 3 zijn de G^2 -waarden gegeven voor de modellen met twee, drie, vier en vijf latente klassen. Alle modellen moeten verworpen worden bij een stricte toetsing met een 5% verworpinggebied. Om na te gaan of dit een gevolg zou kunnen zijn van een specifieke vorm van ons model hebben we de residuen van het minst gestricteerde model dat met vijf latente klassen nader bestudeerd. Hierbij troffen we geen interpreteerbare patronen in de residuen aan, of specifieke uithijters. Daarom nemen we aan dat de significantie van de toedellen een gevolg is van de steekproefgrootte.

De kolom parameters van het latente budgettenmodel $\hat{\pi}_{ikpj}^{BX}$, d.w.z. de waarden van de latente budgetten, zijn voor elk van de modellen te vinden in Tabel 4. Voor het model met twee latente budgetten is het eerste budget het budget waar kinderen een erg lage kans, nl. .011, hebben om drop out te worden, zij gaan niet naar het LBO en minder vaak dan gemiddeld (zoals aangegeven voor de marginale kansens) naar het (M)BO. In plaats hiervan gaan zij vaker naar algemeen voorbereidend onderwijs. In het tweede latente budget zijn zij vaker dan gemiddeld drop out, gaan zij vaker dan gemiddeld naar het LBO en (M)BO, en minder vaak dan gemiddeld naar algemeen voorbereidend onderwijs. Voor het model met drie latente budgetten gaan kinderen in het eerste latente budget (i=1) vooral vaker naar het LBO en worden vaker dan gemiddeld drop out. In budget 2 gaan zij vaker dan gemiddeld naar HAVO en VWO, en in budget 3 gaan zij vaker naar MAVO, HAVO en MBO, maar niet vaker dan gemiddeld naar VWO. In het model met 4 latente budgetten lijkt het derde latente budget erg op het tweede latente budget uit het model met 3 latente budgetten. Vergeleken met het model met drie latente budgetten veranderen de budgetten met hoge kanssen op beroepsonderwijs: een nieuw budget, namelijk het tweede, is gevormd, dat is afgeleid uit het eerste en het derde van het model met drie latente budgetten. Het oont een grote kans om (M)BO te doen, en een ook nog wel grote kans om LBO en MAVO te doen. In het model met vijf latente budgetten is het budget voor VWO nog steeds ongeveer hetzelfde (i=2), en worden opnieuw de budgetten voor beroepsonderwijs verder gedifferentieerd.

Tabel 3. Chi-kwadraat toetsen voor latente budget modellen.

	G^2	df	% afwijking van $K=1$
T=1	4612	415	0.0
T=2	1113	328	75.8
T=3	441	243	90.4
T=4	226	160	95.1
T=5	116	79	97.5

iteerd. Over het geheel genomen vinden we, wanneer we van het model met drie naar het model met vijf latente budgetten gaan, dat er altijd een budget is met een kans groter dan nul om naar het VWO te gaan en een kans van nul om naar het LBO en het (M)BO te gaan. Als het aantal budgetten toeneemt, wordt een meer verfijnde beschrijving gegeven van budgetten die een kans hebben die groter is dan nul om beroepsonderwijs te doen.

Gegeven de steekproefgrootte zijn we tevreden met een beschrijving van drie latente budgetten. Ofschoon dit model significant is, is de afwijking tussen $G^2 = 441$ en $df = 243$ niet enorm, en het model beschrijft $(4612 - 441)/4612 = .904$ van de afwijking van het onafhankelijke model (8). De proportionele winst die wordt gemaakt door te kiezen voor een model met vier latente budgetten is gering. Daarom kiezen we voor het model met drie latente budgetten, en zullen hiervan de parameters nader interpreteren.

We beginnen met de bestudering van de parameters $\hat{\pi}_{ikpj}^{ACFX}$. Deze parameters zijn grafisch weergegeven, voor elke TIB-score p en geslacht i afzonderlijk. Dit levert $7 \times 2 = 14$ plaatjes, zie Figuur 2. In elke plot zijn horizontaal de zes categorieën van sociaal milieu k afgezet, en verticaal is de kans gespecificeerd om naar een van de latente budgetten t te gaan. Elk plaatje heeft 18 punten, want kinderen uit elk van de zes sociale milieus kunnen naar elk van de drie latente budgetten gaan. Punten die bij hetzelfde latente budget horen zijn door een lijn verbonden, zodat elk plaatje drie lijnen heeft. We hebben om de volgende redenen gekozen voor deze wijze om de parameters af te beelden. Ten eerste, als sexe geen invloed zou hebben op de kans om naar een van de latente budgetten te gaan, dan zou het plaatje links gelijk zijn aan het plaatje rechts. Verschillen tussen het linker en het rechter plaatje geven zo dus verschillen weer in de manier waarop jongens en meisjes naar de latente budgetten gaan. Ten tweede, als sociaal milieu geen invloed zou hebben op de kans om naar een van de latente budgetten te gaan, dan zouden alle lijnen horizontaal lopen. Afwijkingen hiervan zijn nu eenvoudig waar te nemen. Het is duidelijk dat de kans om naar een latente budget te gaan in hoge mate beïnvloed zal worden door de TIB-score, omdat de verschillende eindniveaus niet alleen verschillende typen onderwijs representeren (namelijk algemeen voorbereidend vs. beroeps-) maar ook onderwijs van een

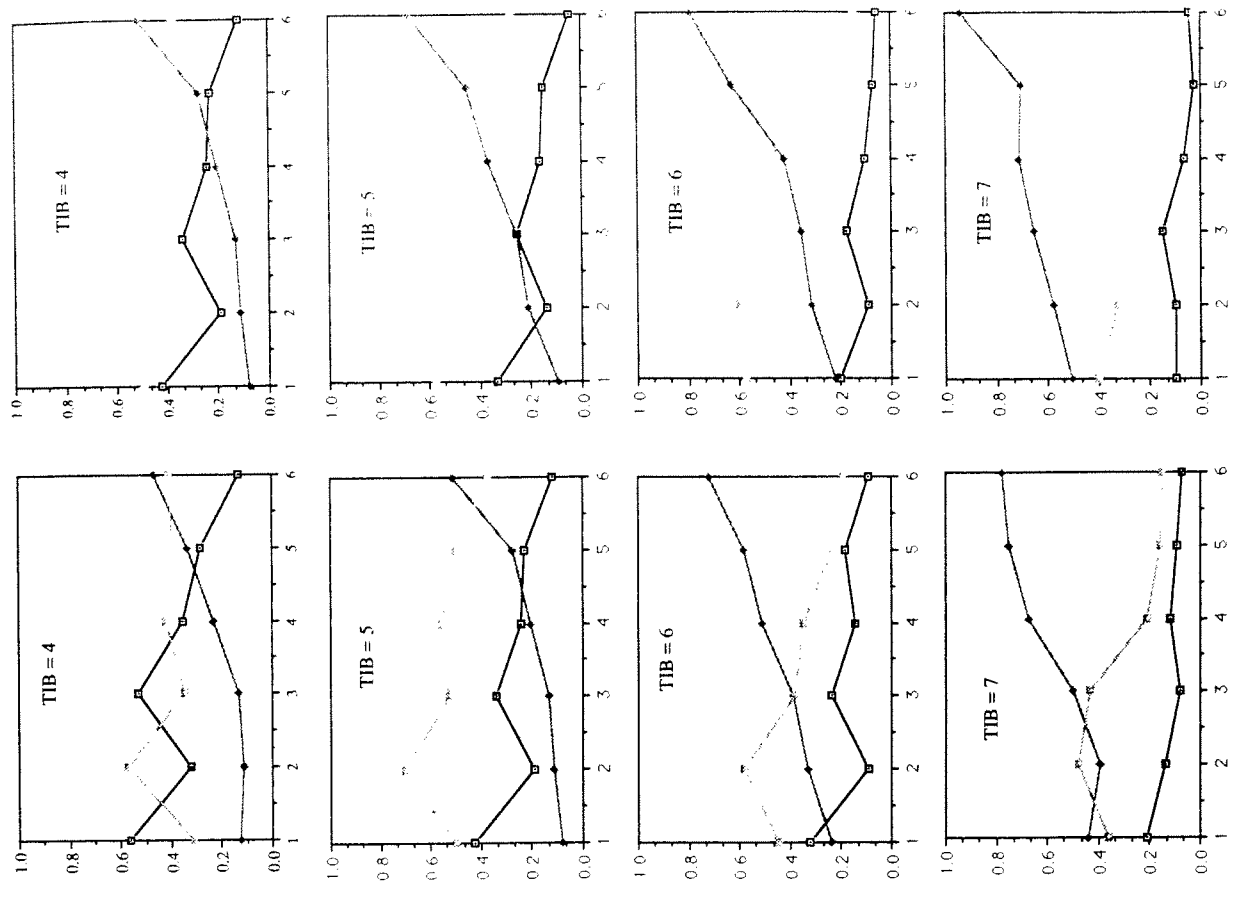


Fig. 2. Representaties van de rijparameters voor het ongerestricteerde model met 3 latente budgetten. Er zijn aparte plaatjes voor elke combinatie van sexe en TIB-score. Een lijn in een plaatje verbindt de kans om naar hetzelfde latente budget te gaan. Horizontaal zijn de niveaus van sociaal milieu uitgezet (1=arbeider, 2=boer, 3=middenstander, 4=lagere employee, 5=middelbare employee, 6=hogere employee), en verticaal zijn de kansen uitgezet om naar elk van de drie latente budgetten te gaan.

rschillende hoogte. Daarom zien we, als we de plotjes van boven (TIB 1) naar beneden (TIB vergelijken, dat het lijntje met de open blokjes steeds lager komt te liggen. Dit is niet r wonderlijk, want deze lijn geeft de kans weer om naar latent budget 1 te gaan, en in dit idget gaat 65.8% van de kinderen naar het LBO en 16.0% wordt drop-out, en dit gebeurt tuurtlijk vaker naarmate de TIB-score lager is.

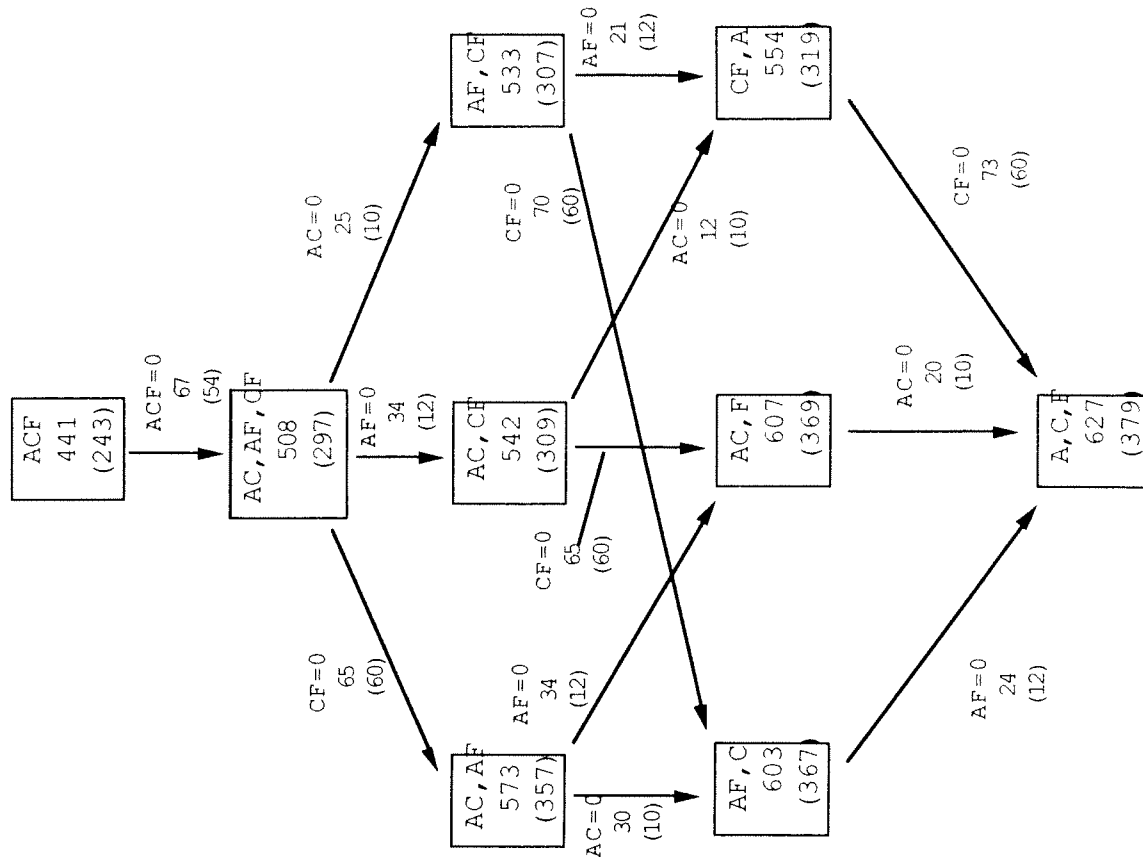
Het is mogelijk verschillende interessante aspecten uit deze plaatjes af te leiden. Om er een ar te noemen: voor alle niveaus van TIB zien we dat kinderen uit hogere sociale milieus (5 en een veel grotere kans hebben om naar het tweede latente budget te gaan, dit is het latente idget voor HAVO (36.7%) en VWO (53.0%). Hun kans om naar het eerste latente budget te an is veel kleiner. Het omgekeerde geldt voor kinderen uit sociale milieu 1 (geschoolede en geschoolede arbeider): gegeven hun TIB-score is hun kans om naar het eerste latente budget te an over het algemeen het grootste. Over het algemeen geldt dat, als de vader van de kinderen er of boerenarbeider is (sociaal milieu 2), dat zij een grotere kans dan gemiddeld hebben om ar latente budget 3 te gaan; in dit budget hebben zij een grotere kans om (M)BO gevolgd te abben. Merk op dat de late budgetten parameters eenvoudig te interpreteren zijn: zij laten niet leen zien dat er iets aan de hand is (bijvoorbeeld: meisjes gaan minder vaak dan jongens naar idget 1), zij laten ook zien hoe sterk de effecten zijn.

We gaan nu de effecten in de plaatjes van Figuur 2 toetsen door de rijparameters π_{ikpt}^{ACFX} : restricties. Dit doen we door gebruik te maken van de factoriële structuur in de rijen, zoals eze gespecificeerd is door de verklarende variabelen A, C en F. De restricties die we introduce- en zijn eenvoudig te begrijpen in termen van de parameters van het loglineaire model voor de itente kansen π_{ikpt} : de parameters λ_{it}^{AX} , λ_{it}^{CX} , λ_{it}^{AFX} , λ_{it}^{ACX} , λ_{it}^{CFX} , λ_{it}^{ACFX} en λ_{it}^{ACFX} . De :stricties hebben een hiërarchische vorm, dat wil zeggen dat als bijvoorbeeld $\lambda_{it}^{AFX} = 0$, dan eldt dat alle hogere orde interacties waarin F en X zitten ook nul zijn. Daarom kunnen we de odellen voor $\log \pi_{ikpt}^{ACFX}$ noteren door de hoogste termen die worden gefit, waarbij we X eglaten, omdat X in elk model zit. We gaan hier niet in op de manier waarop dergelijke istricties kunnen worden opgelegd, maar verwijzen hiervoor naar van der Heijden, Mooijaar n de Leeuw (1990).

In de eerste plaats leggen we de restrictie op dat alle lijnen horizontaal lopen. Dit model oteren we als AF. Het past erg slecht: $G^2 = 2101$. Het is duidelijk dat in de populatie de lijnen iet horizontaal zijn. Een vergelijkbaar model, dat de restricties oplegt dat elk plaatje voor de ongens gelijk is aan elk plaatje voor de meisjes, past veel beter, maar nog steeds vrij slecht: dit model CF (C voor sociaal milieu) heeft een fit van $G^2 = 727$.

Nu leggen we systematisch alle mogelijke restricties op aan de rijparameters, met als meest istrictieve model het model met alleen de hoofdeffecten A, C en F (zie Tabel 5). Het model AC, AF, CF veronderstelt dat er interactie is tussen sexe en sociaal milieu op de latente budgetten, :venzo interactie tussen sexe en TIB, en tussen sociaal milieu en TIB, maar dat er geen interactie ussen sexe, sociaal milieu en TIB gezamenlijk is in hun relatie met de latente budgetten. De fit is rrij goed ($G^2 = 508$, $df = 297$), en het verschil in fit met het ongerestriceerde model is niet :root: $G^2 = 508 - 441 = 67$, $df = 297 - 243 = 54$. Als we naar beneden gaan in Tabel 5, dan zien we dat de fit afneemt naar $G^2=627$ met $df=379$. Vergeleken met het ongerestriceerde model winnen we $379-243=136$ vrijheidsgraden, voor een firverlies van $627-441=186$. Het is duidelijk lat er informatie verloren gaat, maar aan de andere kant wordt de interpretatie veel eenvoudiger, n de parameters worden stabiel. Er is wat informatie verloren gegaan door het op nul stellen van de interactie tussen sociaal milieu en TIB (CF, conditionele toetsen G^2 zijn tussen 65 en 73 voor 60 df) en iets meer door de interactie tussen sexe en sociaal milieu op nul te stellen (AC, :nditionele toetsen zijn tussen 12 en 30 voor 10 df) en door de interactie tussen sexe en TIB op nul te stellen (AF, conditionele toetswaarden liggen tussen 21 en 34 voor 12 df). We gaan nu over tot de interpretatie van het meest restrictieve model A,C,F. Dit model correspondeert met het loglineaire model (vergelijk model (7))

Tabel 5. Modellen met geresliceerde rijparameters voor $T=3$ budgetten. Hierarchical modellen, alleen termen waarin X zit zijn gegeven. A is sexe, C is sociaal milieu en F is TIB. In de blokken staan de modellen, hun fit (likelihood ratio toets G^2) en het aantal vrijheidsgraden, naast de pijlen staan de termen die geresliceerd zijn. Zie verder de tekst.



Tabel 6: latente budget voor T = 3 voor model S.P.T in tabel 5.

	T=3		
	i=1	i=2	i=3
1. Drop out	.177	.025	.005
2. LBO	.701	.038	.006
3. MAVO	.092	.090	.331
4. HAVO	.000	.337	.228
5. VWO	.015	.500	.000
5. (M)BO	.015	.011	.430
Budget grootte	1.000	1.000	1.000
	.304	.274	.422

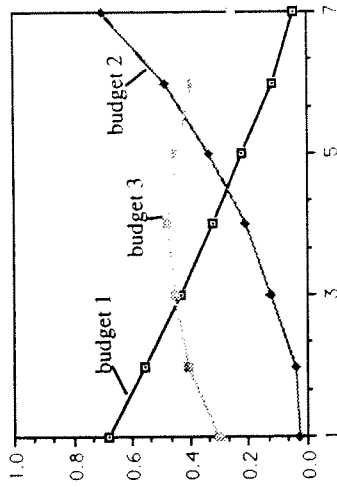
$$\log \pi_{ikpjt} = \lambda + (\lambda_i^A + \lambda_k^C + \lambda_p^F + \lambda_{ik}^{AC} + \lambda_{ip}^{AF} + \lambda_{kp}^{CF} + \lambda_{ikp}^{ACF}) + \lambda_j^B + \lambda_t^X + (\lambda_{it}^{AX} + \lambda_{kt}^{CX} + \lambda_{pt}^{FX}) + \lambda_{jt}^{BX} \quad (9)$$

De schattingen van de parameters van de latente budgetten $\hat{\pi}_{ijt}^{BX}$ voor dit model zijn vergelijkbaar met die voor het ongeïmpliceerde model met drie latente budgetten, vergelijk Tabel 6 met tabel 4. De plaatjes met schattingen $\hat{\pi}_{ikpjt}^{ACFX}$ zijn soms verschillend, zie Figuur 2. Deze plaatjes worden nu het meest eenvoudig bestudeerd door gemiddelde rijparameterschattingen af te leiden als $\hat{\pi}_{it}^{AX} \equiv \hat{\pi}_{i+++} / \hat{\pi}_{i++++}$, $\hat{\pi}_{kt}^{CX} \equiv \hat{\pi}_{k++} / \hat{\pi}_{k+++}$ en $\hat{\pi}_{pt}^{FX} \equiv \hat{\pi}_{p++} / \hat{\pi}_{p+++}$ (vergelijk de wijze waarop de parameters voor het gewone LBA model gerelateerd zijn aan de latente kansen, zie tekst bij formule (5)). Op deze manier verkrijgen we parameters voor TIB alleen, voor sociaal milieu alleen en voor sexe alleen. Plaatjes van deze parameters worden gegeven in Figuur 3. Het plaatje voor TIB-scores laat zien dat de kans om naar latent budget 1 (vnl. LBO, drop out) afneemt als TIB toeneemt, de kans om naar budget 2 (vnl. MAVO, HAVO, MBO) gaat toeneemt als TIB toeneemt, en de kans om naar budget 3 (vnl. MAVO, HAVO, MBO) gaat toeneemt van TIB 1 tot TIB 4, om daarna af te nemen. In het plaatje voor sociaal milieu is de kans om naar budget 1 te gaan kleiner voor kinderen van boeren/boerenarbeiders (2), en middelbare- en hogere employees (5,6), de kans om naar latent budget 2 te gaan neemt snel toe van lagere naar hogere employees, en de kans om naar budget 3 te gaan is iets hoger voor kinderen van boeren/boerenarbeiders en iets lager voor kinderen van hogere employees. In het plaatje voor sexe vinden we dat er geen verschil is voor jongens en meisjes in hun kans om naar budget 1 te gaan. Er is een verschil in hun kans om naar budget 2 en 3 te gaan: voor jongens zijn deze kansen vrijwel gelijk, maar voor meisjes is de kans om naar budget 3 te gaan groter dan die om naar 2 te gaan.

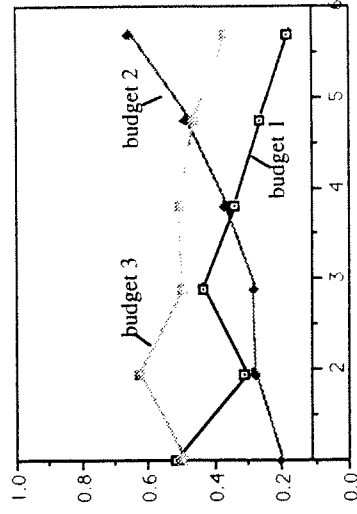
5. CONCLUSIE

Latent budget analyse geeft inzicht in deze data. De interpretatie van LBA als een MIMIC-model toont de kansen die kinderen hebben om in een latent budget terecht te komen. De latente budgetten specificeren de kansen om specifieke eindniveaus te halen. De parameters zijn eenvoudig te interpreteren, zodat de processen die een rol spelen tussen de verklarende variabelen TIB, sexe en sociaal milieu eenvoudig te doorzien zijn. De op te leggen restricties staan een simpelere interpretatie toe.

Effect van TIB - score



Effect Sociaal Milieu



Effect van sexe

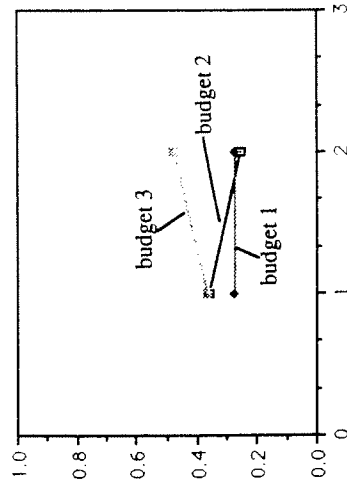


Fig. 3. Hoofdeffecten voor model A, F, C in Tabel 5. Horizontaal zijn de categorieën van elk van de verklarende variabelen uitgezet, en verticaal de kansen om naar elk van de latente budgetten te gaan.

NOOT

Themanummer longitudinaal onderzoek van schoolloopbanen, TOR, 1990, nr. 1 en Themanummer Multivariëtelanalyse, TOR, 1991 of 1992.

LITERATUUR

- Jöreskog, C.C. (1981) Latent structure models of mobility. *American Journal of Sociology*, 86, 836-868.
- de Leeuw, J. en van der Heijden, P.G.M. (1988). The analysis of time-budgets with a latent time-budget model. In: E. Diday et al. (Eds.), *Data analysis and Informatics 5*, Amsterdam: North Holland, p.159-166.
- de Leeuw, J., en van der Heijden, P.G.M. (1991). Reduced rank models for contingency tables. *Biometrika*, 78, 239-232.
- de Leeuw, J., van der Heijden, P.G.M., en Verboon, P. (1990). A latent time budget model. *Statistica Neerlandica*, 44, 1-22.
- Jöreskog, Z. (1979). Singular value decomposition of probability matrices: probabilistic aspects of latent dichotomous variables. *Biometrika*, 66, 339-344.
- Jöreskog, Z. (1983). Latent conditional independence in two-way contingency tables: a diagnostic approach. *British Journal of Mathematical and Statistical Psychology*, 36, 114-122.
- Jöreskog, Z. (1984). On some similarities between canonical correlation models and latent class models for two-way contingency tables. *Biometrika*, 71, 523-529.
- Good, I.J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics*, 11, 823-831.
- Goodman, L.A. (1987). New methods for analyzing the intrinsic character of qualitative variables using cross-classified data. *American Journal of Sociology*, 93, 529-583.
- Grover, R. (1987). Estimation and use of standard errors of latent class model parameters. *Journal of Marketing Research*, 24, 298-304.
- Grover, R. en Srinivasan, V. (1987). A simultaneous approach to market segmentation and to market structuring. *Journal of Marketing Research*, 24, 139-153.
- Haberman, S.J. (1979). *Analysis of qualitative data (2 vols.)*. New York: Academic Press.
- Hagenaars, J.A. (1990). *Categorical longitudinal data. Log-linear, panel, trend and cohort analysis*. London: Sage.
- Luijckx, R. (1987). Loglinear modelling with latent variables: the case of mobility tables. In: W.Saris en I.Gallhofer (Eds.) *Sociometrics Research: Vol.2*. London: MacMillan.
- Marsden, P.V. (1985). Latent structure models for relationally defined social classes. *American Journal of Sociology*, 90, 1002-1021.
- McCutcheon, A.L. (1987). *Latent class analysis*. Beverly Hills: Sage.
- Meester, A. en de Leeuw, J. (1983). *Intelligentie, sociaal milieu en de schoolloopbaan*. Leiden: Vakgroep Datatheorie.
- Meester, A. en de Leeuw, J. (1985). De regressie van intelligentiescores op schoolloopbaanvariabelen onder constanthouding van sociale achtergrond en geslacht. *Tijdschrift voor Onderwijsresearch*, 10, 2-20.
- van de Pol, F. en R. Langeheine (1990). Mixed Markov latent class models: From description towards explanation. In: C.C. Clogg (Ed.) *Sociological methodology 1990*. Oxford: Blackwell.
- van der Heijden, P.G.M., Mooijaart, A. en de Leeuw, J. (1989). Latent budget analysis. In: A. Decarli, B.J. Francis, R.Gilchrist en G.U.H. Seeber (Eds.). *Statistical Modelling*. Proceedings, Trento, 1989. Berlin, Springer Verlag.
- van der Heijden, P.G.M., Mooijaart, A. en de Leeuw, J. (1992). *Constrained latent budget analysis*. Utrecht: University of Utrecht, ISOR, Series Methods, MS 90-1. Geaccepteerd door Sociological Methodology 1992, gededit door P. Marsden.
- van der Heijden, P.G.M., de Leeuw, J. en Mooijaart, A. (1990). On the relation between latent class analysis and correspondence analysis. In: *Proceedings of Fifth International Workshop on Statistical Modelling*. Toulouse: Université Paul Sabatier.

Manuscript ontvangen 14-5-1991

Definitieve versie ontvangen 30-8-1991

APPENDIX: DE SMVO DATA

School types: 1=drop out, 2=LBO, 3=MAVO, 4=HAVO, 5=VWO, 6=(M)BO. Sociaal Milieu (SES) 1=geschoolde en ongeschoolde arbeiders 2=boeren en boeren arbeiders, 3=middenstanders, 4=lagere employeés, 5=middelebare employeés, 6=hogere employeés. TTB scores: aantal items correct beantwoord zijn 1=1-14, 2=15-17, 3=18-20, 4=21-23, 5=24-26, 6=27-29, 7=30-33

School type	Jongens					Meisjes						
	1	2	3	4	5	6	1	2	3	4	5	6
SES 1 TTB 1	43	126	23	5	2	17	28	87	24	13	3	35
2	41	172	58	20	9	28	29	131	57	15	0	74
3	50	271	83	58	24	87	67	209	128	59	6	141
4	64	268	131	93	44	111	64	200	157	95	34	194
5	43	202	121	113	47	109	35	163	177	105	39	201
6	11	78	60	62	43	78	20	54	106	92	48	103
7	4	15	20	23	27	19	2	10	22	40	38	28
SES 2 TTB 1	3	13	1	1	1	8	2	8	5	1	0	5
2	3	18	9	0	0	10	2	14	10	4	0	12
3	2	18	12	15	3	23	5	18	16	19	3	26
4	8	25	15	14	9	47	0	18	23	21	8	46
5	5	25	16	12	16	35	0	13	28	21	15	39
6	2	4	7	20	11	22	5	6	19	37	15	30
7	0	3	2	5	7	9	0	4	4	12	17	10
SES 3 TTB 1	11	17	6	1	1	10	7	12	11	2	0	8
2	9	37	11	6	2	10	6	29	11	5	1	11
3	23	59	26	12	6	29	16	43	30	19	4	38
4	12	72	34	23	14	38	18	39	39	36	13	49
5	11	40	26	37	25	36	16	32	54	54	25	39
6	7	20	26	25	30	25	11	12	28	41	20	24
7	3	1	7	9	12	9	2	3	3	16	7	3
SES 4 TTB 1	9	29	13	4	1	4	3	15	6	3	0	10
2	9	38	21	5	4	13	10	24	26	7	2	29
3	12	56	47	37	15	27	12	54	40	37	15	35
4	11	62	52	54	26	43	15	39	64	56	27	61
5	12	48	62	55	37	30	9	31	54	87	44	52
6	6	15	33	40	45	24	7	11	35	49	39	39
7	3	4	7	17	23	7	2	3	5	23	26	9
SES 5 TTB 1	5	25	14	9	3	9	6	20	8	3	1	12
2	8	26	30	23	7	11	9	22	24	19	4	30
3	13	60	65	39	35	50	10	42	50	44	33	59
4	20	79	91	94	71	70	17	58	97	82	55	79

5	11	58	70	95	63	11	44	89	103	101	70	
6	9	39	44	71	107	40	5	17	46	117	104	47
7	4	7	9	28	57	12	2	3	28	49	70	21

Boekbesprekingen

Dr. C.A.W. Glas, *Contributions to Estimating and Testing Rasch-models*. Dissertatie, Universiteit Twente, 1989.

Appendix (vervolg)

ASES 6	TIB 1	4	6	10	6	4	3	5	2	6	1	1	5
2	7	14	15	11	5	12	4	3	6	18	2	11	
3	5	31	34	39	21	23	5	16	24	33	16	21	
4	10	16	45	54	52	36	9	16	44	83	46	29	
5	7	16	44	71	105	28	7	7	40	80	83	27	
6	3	12	24	40	85	19	8	7	32	66	100	15	
7	3	4	9	16	52	9	1	3	10	29	51	1	

Sinds het einde van de zeventiger jaren bestaat in Nederland een snel gegroeide groep van psychometrici die zich intensief en grensverleggend bezig houden met latente trek (of item-respons) theorie en in het bijzonder met het Rasch-model, zijn varianten en zijn toepassingen. Het is verheugend dat deze groep ook internationaal sterk de aandacht trekt (zie Hambleton en Swaminathan, 1985, in een themanummer gewijd aan Testtheorie), en dat Paul Holland (1990) een theorema naar Nederland heeft vernoemd: "The Dutch Identity". Een groot aantal publicaties, o.a. in Psychometrika, en dissertaties getuigen van de belangrijke Nederlandse bijdragen aan de ontwikkeling van de item-respons theorie (IRT). Het CITO, waar de auteur van het hier besproken proefschrift sinds 1983 werkzaam is, heeft daarin een wezenlijk aandeel.

De dissertatie van Glas (gepromoveerd bij prof. W. van der Linden en Prof. A. Albers, met Dr. Verhelst als co-promotor), waarvan enkele belangrijke onderdelen al eerder waren verschenen (Glas, 1988a, 1988b; Glas en Verhelst, 1989), is een der meest indrukwekkende recente bijdragen aan de item-response theorie. (Het is verleidelijk, maar zou de strekking van deze bespreking te buiten gaan, om hier ook andere dan die van Glas te noemen).

Het Rasch-model is een IRT-model van verbluffende eenvoud: het beschrijft de waarschijnlijkheid van een positieve responsie op een item als een logistische functie van het verschil tussen een persoonsparameter ('vaardigheid') en een itemparameter ('moeilijkheid'). Het heeft een verrassende methodologische betekenis vanwege het principe van 'specifieke objectiviteit', dat het belichaamt. Juist wat de preciese inhoud van dit principe betreft, zijn de geleerden het nog niet eens, en ook Glas heeft er mee geworsteld. Ik kan met zijn behandeling ervan niet gelukkig zijn, en ik deel zeker niet zijn oordeel (p. 12) dat ook het twee-parameter model van Birnbaum aan 'specifieke objectiviteit' zou voldoen. In ieder geval is men het er wel over eens dat het Rasch-model toestaat uitspraken over itemmoeilijkheden te doen die onafhankelijk zijn van de steekproef van subjecten, en omgekeerd, dat uitspraken over personen mogelijk zijn onafhankelijk van de steekproef van items uit een universum van items. Daardoor opent het Rasch-model belangrijke perspectieven voor, o.a., praktische problemen zoals item-banking, test-equivalering, item-bias (differential item functioning) en vergelijking van verschillende groepen van subjecten. Deze perspectieven bestaan uiteraard slechts indien aan de eigenschappen van de parameters en het toetsen van 'goodness of fit' op (Het is jammer dat we nog steeds niet beschikken over bevredigende Nederlandse termen). Glas' werk heeft daaraan fundamentele nieuwe bijdragen geleverd. Er zijn in de loop van verscheidene decennia diverse technieken voor het schatten en toetsen ontwikkeld, en het is gebleken dat, vooral wat de toetsing betreft, de interpretatie van de toetsingsgrootheid niet eenduidig is: verschillende toetsen zijn gevoelig voor verschillende aspecten van de datastructuur. Van den Wollenberg (1982) was, meen ik, de eerste die daarop uitdrukkelijk de aandacht vestigde. In de loop der tijd zijn verscheidene toetsen ontwikkeld, die grofweg in twee categorieën vallen: likelihood-ratio toetsen, en Pearson χ^2 -achtige toetsen. Binnen deze laatste zijn te onderscheiden: (1) toetsen gebaseerd op de frequentie van positieve responsies op een item, i , bij subjecten met een ruwe score r , en (2) toetsen gebaseerd op de frequentie van positieve responsies op twee items, i en j , bij subjecten met een ruwe score r . Van de steekproefverdeling van deze toetsen was niet al te veel bekend, en dat kwam ook omdat de eigenschappen van deze toetsen mede afhangen van de wijze waarop de parameters worden geschat. Voor die parameterschatting bestaan verschillende methoden. Het meest direct aansluitend op de eigenschappen van het Rasch-model is de Conditionele Maximum Likelihood schatting (CML). Daarnaast bestaat de Unconditional ML-schatting, en vooral sinds de ontwikkeling van het EM-algoritme, de Marginale ML-schatting (MML). Deze laatste