



HET SCHATTEN VAN SCHOOLEFFECTEN

Problemen bij de keuze van het model en de techniek.

Ita Kreft
Vakgroep AVO-UvA
Prinsengracht 227
1015 DT Amsterdam

Jan de Leeuw
Vakgroep Datatheorie FSW-RUL
Middelstegracht 4
2312 TW Leiden

uitgave

**SAMENWERKINGSVERBAND
ONDERWIJSKUNDE
RANDSTAD**

nr. 7

Samenvatting

In een recent artikel bespreken Aitkin en Longford een aantal varianten van het lineaire model die gebruikt kunnen worden om schooleffecten op schoolloopbaanvariabelen te kwantificeren. Ze laten daarbij zien dat schooleffect alleen gedefinieerd kan worden binnen een volledig gespecificeerd statistisch model, en dat sommige van de onderzochte modellen bruikbaar lijken dan anderen. Aitkin en Longford spreken hun voorkeur uit voor *random effect modellen*, ook wel *variantie component modellen* genoemd. De door hen bestudeerde random effect modellen zijn een bijzonder geval van de onder andere door De Leeuw en Kreft bestudeerde *random coëfficiënt modellen*. In dit artikel bespreken we, aan de hand van een SMVO-deelbestand, hoe in random coëfficiënt modellen schooleffecten gedefinieerd kunnen worden. We laten zien hoe de kwantificatie van het schooleffect verandert, wanneer het model restrictiever gemaakt wordt, of in andere zin gemodificeerd wordt. De praktische consequenties voor schooleffect- en schoolloopbaanonderzoek worden besproken. De doelstelling van onze discussie is hier om aan te tonen dat sommige modellen eenvoudiger zijn en tevens beter bij de gegevens passen. Dit laatste wordt uitgedrukt in een chi kwadraat toets. De verschillende modellen die we op ons voorbeeld zullen fitten zijn het covariantie model, waarvan we concluderen dat het goed op de gegevens past, maar niet eenvoudig is vanwege het grote aantal parameters dat geschat moet worden. Het tweede model is het contextuele model, dat wel eenvoudig is, maar niet goed op onze data past. Het derde en laatste model is het random coëfficiënt model. De aanname van gecorreleerde errortermen op het individuele niveau binnen eenzelfde context maakt het model goed passend bij onze gegevens. Het is dit model, dat wij verkiezen, omdat het goed bij de data past, maar ook omdat het eenvoudig is.

Trefwoorden

multilevel analyse, schoolkenmerken, schoolloopbanen, modelspecificatie, lineaire modellen

1 Theoretisch kader.

Inleiding. In dit paper houden we ons bezig met de karakteristieken van onderzoek naar schooleffecten op schoolloopbanen, en de implicaties daarvan voor de statistische modellen. Dit betekent, dat we niet op de theoretische en methodologische redenen voor de keuze van een multi-niveau benadering zullen ingaan. Hiervoor verwijzen we naar Boyd and Iverson (1979) en Blalock (1984). We hebben het daarom alleen over de statistische aspecten van de verschillende modellen, de rol van de theorie en van de juiste keuze van een model.

De definitie van het schooleffect door middel van het model. Het vaststellen van schooleffecten op schoolloopbanen kunnen we beschrijven als: hoe is de relatie tussen de uitkomsten van onderwijs, schoolprocessen en eigenschappen van leerlingen. Een belangrijk onderdeel van dit proces is de verklaring van variabiliteit van genoemde relaties tussen scholen. Een veel gebruikt analysemodel hierbij is multi-pele regressie, waarbij de uitkomsten van het onderwijs (de 'output') gerelateerd worden aan aanvangskenmerken van de leerlingen en aan schoolvariabelen (de 'input'). Hoewel de meeste sociologen het er over eens zullen zijn, dat contexten het gedrag van individuen beïnvloeden, is er op contextuele data analyse veel kritiek geleverd. Zie Swanborn (1973, 1978) en Hauser (1974). De discussie wordt echter bemoeilijkt omdat de gebruikte termen slecht gedefinieerd worden (De Leeuw en Kreft, 1984), en ook omdat de discussie zich soms op verschillende niveaus beweegt, zonder dat de discussianten zich hiervan bewust zijn. Met name het gebruik van een analysetechniek en het gebruik van een analysemodel worden door elkaar gehaald. In dit paper doen we een poging dit te voorkomen door in deze paragraaf te definiëren wat we bedoelen als we het hebben over een contextueel- of schooleffect.

Binnen het multiple regressie model zijn veel variaties mogelijk. Deze variaties worden in onderzoek ook door elkaar gebruikt, met als gevolg dat de belangrijkheid van het schooleffect zeer verschillend is beoordeeld. Schooleffecten worden gedefinieerd met behulp van een gekozen model. De techniek wordt toegepast, nadat de

veronderstellingen over het model gemaakt zijn. De keuze van een techniek is aan het model ondergeschikt, en in feite in gelijke mate irrelevant als de keuze of een correlatiecoëfficiënt berekend wordt met SPSS op een computer of met de hand op een rekenmachine. Een computer programma heeft binnen deze opzet geen andere functie dan het schatten van de vrije parameters en het toetsen van de veronderstellingen, die gemaakt zijn in het model. Hierna vindt een eventuele aanpassing van het model plaats op basis van de *goodness of fit* en kunnen eventueel weer nieuwe keuzen gemaakt, of nieuwe computer programma's aangesproken worden. Al deze keuzen hebben gevolgen voor de definitie van het geschatte schooleffect. Dus een ander model, met andere specificaties, zal een andere schooleffectmaat opleveren. We zien dit geïllustreerd in Aitkin en Longford (1986)

De techniek is hierbij het hulpmiddel voor de onderzoekster waarmee gekeken wordt, of het door haar aangenomen model op de gegevens past, waarna een volgende cyclus kan volgen van modelaanpassing en modeltoetsing. In feite vervangen deze cycli van modelopstelling en modelaanpassing het oorspronkelijke experiment uit de natuurwetenschappen. We kunnen daarom zeggen dat 'modellen fitten' het antwoord is van de sociale wetenschap op het ontbreken van mogelijkheden tot sociale experimenten. In de latere paragrafen zullen we de bovenstaande opvattingen in de praktijk toetsen, en aantonen welke invloed de verschillende modelspecificaties hebben op de grootte van het gemeten schooleffect.

We gaan hierbij in op de door Tacq (1986) gestarte discussie over de relatie tussen probleemstelling en analysetechniek. Onze discussie gaat echter verder, met name omdat we de keuze van de techniek zien als een keuze die afhankelijk is van meer zaken dan de probleemstelling alleen. Aan de hand van de probleemstelling, de kennis van het terrein waarop het onderzoek plaats vindt, en het domein waaruit de gegevens komen, stellen we een model op. De toetsing van de veronderstellingen van dat model gebeurt met behulp van een techniek. De analyse-techniek verschaft ons de parameters van het model op basis van de onderzoeksgegevens, alsmede een maat voor de juistheid van het vooraf vastgelegde model. De vraag of een model past op de gegevens wordt via statistische toetsing beantwoord. Hierbij geldt de regel dat, bij gelijke passing, een eenvoudig model met weinig parameters te verkiezen is boven een model met veel

parameters. Eenvoudige modellen zijn informatiever en stabiel, ingewikkelde modellen daarentegen moeilijker interpreteerbaar.

Het analyse-niveau. Aanvankelijk werd het probleem van de analyse van meer-niveau gegevens gezien als het probleem van een keuze van het juiste analyse-niveau, in plaats van de keuze voor het juiste model. Dit niveau zou in principe bepaald moeten worden door de vraagstelling, maar werd in de praktijk vaak bepaald door omstandigheden zoals het beschikbaar zijn van gegevens van een bepaald niveau. Dit laatste speelt een grote rol bij de analyse van economische gegevens, omdat men hier vaak alleen gegevens op groepsniveau ter beschikking heeft. Het is dan ook niet verwonderlijk dat vanuit deze traditie het macro-niveau (de school) als de belangrijkste analyse-eenheid wordt gezien.

Vanuit de onderwijspsychologie komt een tweede traditie, die het micro niveau (de leerling) als de belangrijkste en soms als de enig juiste analyse-eenheid beschouwt. Hauser (1970, 1974) is daar de bekendste vertegenwoordiger van. De laatste verdedigt dit standpunt door te stellen, dat groepseffecten artefacten zijn van falende onderzoekstheorieën. Verklaarde varianties vanuit context variabelen zijn meestal het gevolg van onvoldoende pogingen om de invloeden te verklaren vanuit het individu. In feite verwijt Hauser de onderzoeker hier, dat deze probeert het tekort schieten van een theorie op individueel niveau te maskeren door deze aan te vullen met "armzalige" maten voor onderliggende sociale processen, zoals bijvoorbeeld een gemiddeld IQ. In principe komt de opvatting van Hauser er op neer, dat context-variabelen nooit gebruikt kunnen worden, omdat altijd de mogelijkheid blijft bestaan dat de verklaarde variantie toe te schrijven valt aan een niet in het model opgenomen individuele variabele, waardoor en waarlangs de werking van de context-variabele verklaard kan worden.

Beide tradities, uit de economie en uit de psychologie, verenigen zich in de onderzoekaankpak van de contextuele analyse, zoals deze nu in ontwikkeling is. Het is daarom dat contextuele analyse soms beschouwd wordt als een synthese van de Amerikaanse micro- en de Europese macrotraditie in onderzoek (Boyd en Iversen 1979, p. 22).

Het contextuele model. Zowel vanuit de micro- als vanuit de macro-traditie is er veel kritiek gekomen op elkaars werkwijze. Het bekendst is de kritiek van Robinson (1950) op het maken van "crosslevel inferenties", de tendentie, om correlaties gevonden op het ene niveau te interpreteren als houdbaar en geldig voor een lager, of hoger niveau. Robinson toonde aan, dat een correlatiecoëfficiënt van .20 gevonden op het individuele niveau werd opgeblazen tot .95 wanneer de eenheden van de analyse een hogere graad van aggregatie bereikten. Dit effect wordt sindsdien het Robinson-effect genoemd. Bekend onder de naam "ecological fallacy" is de omgekeerde situatie. Als een verband, gevonden op het macroniveau, van toepassing wordt geacht op een lager niveau. Soms zit men er dan naast, omdat niet zonder meer kan worden aangenomen, dat eenzelfde relatie ook aanwezig zal zijn op het microniveau. Een aansprekend voorbeeld is de relatie tussen het stemmen op de Centrum Partij en het voorkomen van veel buitenlanders in stadswijken. Het is duidelijk, dat buitenlanders niet op de Centrum Partij zullen stemmen.

Als verklaring voor deze discrepantie tussen de verbanden van dezelfde variabelen op micro- en macro- niveau wordt wel gegeven, dat variabelen, die zijn afgeleid uit individuele variabelen, zoals groepsgemiddelden, vaak nog wel dezelfde naam hebben maar niet meer dezelfde betekenis. Een verschijnsel dat bekend staat onder de naam de "aggregatiefout" (Hannan, 1970). Of in de woorden van Galtung (geciteerd in Hannan 1970): "Here the fallacy would consist in believing that, because two behavioral elements are found within the same individual, they are also found in the same behavioral, temporal, spatial or mental contexts, we may choose to subdivide the individual into." (1967, p. 46). Beide fouten, de laatst genoemde aggregatiefout en het Robinson-effect, geven aan dat voorzichtigheid geboden is bij het maken van "cross level inferenties."

Als we de bezwaren tegen analyses op één enkel niveau samenvatten, dan komen we tot de conclusie dat aggregatie en analyse op een hoger niveau, maar ook disaggregatie en analyse op een lager niveau, tot problemen kunnen leiden als we uitspraken willen doen die geldig zijn op beide niveaus. Alle pogingen om variabelen van verschillend niveau te herleiden tot één enkel niveau hebben echter gevolgen voor de betrouwbaarheid van de resultaten van de analyse. Aggregatie naar het schoolniveau

elimineert de binnen-scholen variantie van alle individueel gemeten variabelen, waarbij het gevaar van het "Robinson effect" optreedt, een opgeblazen correlatie coëfficiënt dus, en het gevaar van de aggregatiefout. Als we disaggregeren naar het leerlingniveau zitten we weer met andere problemen. Dan krijgen leerlingen op eenzelfde school gelijke scores voor school- of leerkrachtkenmerken, waardoor deze laatste variabelen een binnen-scholen variantie van nul hebben.

In het contextuele model kunnen beide "fallacies" vermeden worden, omdat met meerdere niveaus tegelijk rekening wordt gehouden. Het door ons voorgestelde model is een uitbreiding van de algemeen gebruikte modellen voor lineaire regressie en variantie analyse, waarin verschillende niveaus kunnen worden opgenomen. Het fundamentele idee van deze familie van multilevel lineaire modellen is in principe eenvoudig. Er is sprake van een micromodel, dat apart gedefinieerd wordt voor elke macro-eenheid. In het macromodel worden de parameters van het micromodel gerelateerd aan schoolkenmerken. In ons onderzoek betekent dit, dat elke school zijn eigen model heeft. Schoolkenmerken kunnen globale kenmerken zijn of centrale tendenties van leerlingkenmerken. In ons onderzoek hebben we slechts de beschikking over het laatste soort kenmerken.

2. Hierarchische modellen.

Bij het meten van schooleffecten op schoolloopbanen komen enkele methodologische en statistische problemen voor, waar tot nu toe geen bevredigende oplossingen voor waren. Het eerste probleem is de hierarchische indeling van leerlingen in klassen, van klassen in scholen, en van scholen in regio's. Veel selectie en recruiteringsprocessen liggen aan de toewijzing van leerlingen aan groepen ten grondslag. Het zijn deze selectie processen, die in het verleden hebben geleid tot discussies over de "confounding effects" van leerling- en schoolkenmerken. Of in de woorden van Pedhazur (1975, p. 255): "The overlap between peer quality and schoolquality is so large that there seems hardly any point in referring to them as different factors." Hieruit zouden we een pleidooi kunnen afleiden, die lijkt op het eerder aangehaalde standpunt van Hauser, om context-invloeden te vertalen als processen van interacterende

individueën. Ons onderzoek heeft echter niet deze meer fundamentele theoretische discussie op het oog, als we ons als doel stellen na te gaan wat deze hiërarchische *nesting* betekent voor de keuze van het lineaire model.

De laatste ontwikkelingen op het gebied van de statistiek in schooleffectstudies suggereren, dat variantie component modellen en random coëfficiënt versies van het gebruikelijke lineaire model effectieve methoden bieden voor de geschetste multilevel problemen. We verwijzen naar Burstein e.a. (1978), Tate en Wongbundit (1983), De Leeuw en Kreft (1986), en Aitkin en Longford (1986). Met name maken deze nieuw ontwikkelde modellen het mogelijk om schooleffecten van individuele effecten te scheiden, en daarbij te corrigeren voor aanvangsverschillen in de populatie. Dit betekent, dat enkele bezwaren tegen de wat oudere vormen van contextuele analyse, zoals we die vinden in Boyd and Iversen (1979), worden ondervangen.

3. De twee-stapsprocedure.

Voordat we beginnen met de bespreking van de door ons voorgestelde random coëfficiënt modellen, en de verbeteringen die dergelijke modellen betekenen, enkele opmerkingen over de *twee-stapsprocedure*. Deze is in Nederland geïntroduceerd door Van den Eeden en Saris (1984), en heeft daarna tot enige verwarring aanleiding gegeven. De procedure moet niet gezien worden als een techniek op basis van een duidelijk gespecificeerd model. Deze indruk zou kunnen ontstaan naar aanleiding van de inhoudsopgave van het nuttige overzichtsartikel van Tacq (1986), waarin deze procedure als zodanig wordt voorgesteld. Ook uitspraken als: "aangezien de analyse technieken.....niet veel verder staan dan de tweestapsprocedure..." (l.c., p.2), wijzen in deze richting.

Van de Eeden en Saris (1984) omschrijven zelf de procedure als volgt: "Wij suggereren een procedure bestaande uit twee stappen. In de eerste stap kiest men de elementen per context als eenheden en analyseert men de relaties tussen de variabelen voor deze eenheden per context. Dit betekent dat men evenveel analyses doet als er contexten zijn." (l.c., p.166). Even verderop vervolgen de auteurs: "Hierbij krijgt men voor elke context waarden voor de coëfficiënten.....waarvoor men kan toetsen of de

context invloed heeft door na te gaan of de coëfficiënten van context tot context verschillen." (l.c., p. 166). Maar er is een zekere spanning tussen het eerste en het tweede citaat. Het eerste citaat omschrijft een *techniek*, die uitgevoerd kan worden zonder dat een bepaald model verondersteld wordt. Daar staat tegenover dat het toetsen, waarover in het tweede citaat wordt gesproken, een *model* vereist, omdat de toets alleen afgeleid kan worden op basis van een model. Een model met vooronderstellingen over bijvoorbeeld ongecorrleerde fouten en onafhankelijke normaal verdeelde storingsen. De volledige twee-stapsprocedure, inclusief toetsing, is dus niet toepasbaar zonder het opstellen van een model met de bijbehorende veronderstellingen over storingstermen. Dit betekent, dat binnen het recept van de twee-stapsprocedure er een aantal modelkeuzen overblijven. Een van de belangrijkste is de keuze tussen een random- of een fixed-coëfficiënt model. Van den Eeden en Saris schrijven een random coëfficiënt model (de coëfficiënten uit de eerste stap zijn random variabelen in de tweede stap, nl. gemeten met een storingsterm), maar verbinden daar geen consequenties aan voor de storingstermen, zodat bij hen niet duidelijk wordt welk model verondersteld wordt.

Op basis hiervan is het ook verstandig de "twee-stapsprocedure" te onderscheiden van "een berekening in twee stappen". In het eerste wordt een model aangenomen, terwijl in het laatste alleen wordt gezegd hoe de procedure verloopt, nl. in twee stappen. Dit laatste is een handige en gemakkelijk te beschrijven methode om aan te geven hoe je iets uit kan rekenen als je te maken hebt met gegevens van verschillend niveau. Bij een "twee-stapsprocedure" wordt echter uitgegaan van een vooraf geëxpliciteerd model, waarin voor beide niveaus aannamen voorkomen over de manier waarop de gegevens tot stand zijn gekomen. Een dergelijke twee-stapsmodellering wordt gebruikt, als men er van overtuigd is, dat twee-stapsschatters beter zullen zijn dan één-stapsschatters. Of dit zo is zal in belangrijke mate afhangen van de nauwkeurigheid en de correctheid waarmee het model op ieder van de niveaus is gespecificeerd. Het blijft echter bij het gebruik van de twee-stapsprocedure toch zaak, om direkt, bij de eerste stap al een geschikt model met een daarbij horende geschikte techniek te kiezen, omdat keuzen in de eerste stap in volgende stappen doorwerken. Ieder van de afzonderlijke stappen heeft daarbij betrekking op slechts een enkel niveau, zodat we bij de afzonderlijke stappen niet van een multilevel techniek kunnen spreken. Door combinatie

van de stappen achteraf kunnen we wel multilevel uitspraken doen. We citeren De Leeuw en Kreft (1986, p. 69): "It is clear by now that the terminology single-step and two-step is quite misleading".

Een andere bron van verwarring is het onderscheid tussen de "single- and separate equation", termen die we bij Boyd and Iversen (1979) tegen komen. Vaak worden deze termen verward met de termen "één-staps- en twee-stapsprocedure". En dit geheel ten onrechte, omdat beide laatst genoemde procedures als technieken gebruikt kunnen worden om een single-, maar ook om een separate-equation te berekenen. Een volgende verwarring kan gemakkelijk ontstaan, omdat er ook een "single- and separate-equation *procedure*" bestaat (Tate and Wongbudit, 1983). Onder de single equation procedure wordt een enkele multipiele regressie analyse verstaan, waarbij variabelen van verschillende niveaus in één enkele analyse worden berekend. Voor verdere discussie op dit punt verwijzen we hier naar De Leeuw en Kreft (1986), met name de paginaas 69 en 70.

4. Beperkingen van covariantie analyse modellen met vaste effecten.

Bij een keuze voor een van de vele mogelijke randomcoëfficiënt modellen spelen een groot aantal overwegingen een rol. Sommige zijn gebaseerd op kennis van het onderzoeksterrein, andere op a-priori overwegingen van statistische aard. We geven als voorbeeld de keuze voor een speciaal (restrictief) geval uit deze familie van modellen: het bekende covariantie analyse model. Dit model is "fixed" in de parameters, dwz. dat de parameters worden opgevat als constanten, en niet (zoals in het random model) als variabelen met een storingsterm.

De keuze voor dit vaste effectenmodel is gebaseerd op twee veronderstellingen. Ten eerste wordt verondersteld dat de scholen uit het onderzoek deelpopulaties zijn, en geen steekproef uit een populatie van scholen. De geschatte effecten hebben daardoor alleen betrekking op deze scholen en zijn in principe niet generaliseerbaar naar scholen die niet in het onderzoek zijn opgenomen. In experimenten betekent iets dergelijks, dat alle mogelijke variaties van de onafhankelijke variabele waarin we geïnteresseerd zijn in het onderzoek voorkomen. We willen bijvoorbeeld uitsluitend uitspraken doen over deze

drie meststoffen, of over deze vier leermachines. In veel onderzoek van onderwijs moet echter aangenomen worden, dat de in het onderzoek aanwezige scholen een steekproef zijn uit de mogelijke scholen, waarover je wel generaliserende uitspaken wilt doen. In een dergelijke situatie is het dan ook logisch uit te gaan van "random" effecten in plaats van "fixed" effecten (zie verder Hays, 1974, hoofdstuk 12 en 13). Een andere beperkende veronderstelling van het covariantie analyse model is de aanname, dat de regressiecoëfficiënten (de "slopes") binnen contexten niet verschillen.

Het is duidelijk, dat een keuze voor dit bepaalde vaste effecten model in schooleffect onderzoek niet in alle situaties juist is. Ten eerste gaat het er wel om generalisaties te maken naar de Nederlandse situatie in het algemeen, terwijl er bovendien vanuit verschillende theorien verondersteld wordt, dat een bepaald schoolbeleid van invloed is op de regressiecoëfficiënt. De aanname van gelijke "slopes" binnen contexten verhindert de onderzoeker om de invloed van schoolbeleid op de regressiecoëfficiënt te schatten. Met name bij de evaluatie van het voorrangbeleid zal men geïnteresseerd zijn in schoolkenmerken, die invloed uitoefenen op bijvoorbeeld de relatie tussen het ouderlijk milieu en de schoolprestatie.

Een derde veronderstelling van het vaste effecten model, die al of niet impliciet is bij de keuze voor covariantie analyse, is een aanname over de storingstermen van het individuele niveau. De storingstermen worden immers geacht de invloed samen te vatten van een groot aantal niet gemeten variabelen, die echter wel invloed uitoefenen op de afhankelijke variabelen, en van meetfouten. Van storingstermen wordt aangenomen dat ze random zijn, dwz. niet onderling gecorreleerd. In onderzoek naar schooleffecten ligt het echter voor de hand te veronderstellen, dat een gedeelte van deze ongemeten variabelen binnen een school een zelfde effect zullen hebben, nl. het schooleffect. Hieruit volgt dat veronderstelling van gecorreleerde storingstermen binnen scholen eveneens voor de hand ligt. Deze gecorreleerde storingstermen komen te voorschijn in de random coëfficiënt versies van contextuele analyse, die we hierna bespreken.

5. Random coëfficiënt modellen.

Random of fixed. Het effect van de school wordt binnen een niet-experimentele situatie gedefinieerd door het gehanteerde model en de veronderstellingen ervan. Een van de veronderstellingen van de random coëfficiënt modellen is de aanname met betrekking tot de storingstermen van het tweede niveau. Een uitgebreide discussie over het principiële verschil tussen "fixed" en "random" modellen vinden we in het eerder aangehaalde hoofdstuk 12 uit Hays (1974). Het verschil tussen beide modellen ligt in de veronderstellingen van het onderzoek met betrekking tot de representativiteit van de verschillende contexten. In de woorden van Hays: "When only a subset of the treatments of interest appear in any given repetition of the experiment, the fixed effects model no longer applies." (1974, p. 526).

Het in dit paper gebruikte basismodel is

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \epsilon_{ij}, \quad (1a)$$

met

$$\alpha_j = \alpha_j + u_j, \quad (1b)$$

$$\beta_j = \beta_j + t_j. \quad (1c)$$

Met onderstreping van de variabelen en coëfficiënten in deze vergelijkingen geven we hier en in het vervolg van dit paper aan dat deze random zijn. Model (1) heeft slechts één onafhankelijke variabele op individueel niveau. De x_{ij} zijn de metingen op die variabele (van individu i op school j). In (1b) zien we een storingsterm u_j , die evenals de storingstermen van het individuele niveau ϵ_{ij} alle ongemeten variabelen vertegenwoordigt, maar nu van het contextuele niveau. In de tweede vergelijking hebben we het random intercept $\alpha_j = \alpha_j + u_j$ gedefinieerd. Op dezelfde manier is t_j de

storingsterm van de helling, en definieert (1c) de random slope $\beta_j = \beta_j + t_j$. De structuur van het random-coëfficiënt model is hiermee aangegeven. Een structuur, die op een natuurlijke wijze voortkomt uit de manier waarop de gegevens in de werkelijkheid gegroepeerd zijn. We onderstrepen dit nogmaals met een citaat van Tate en Wongbundit. "We agree with the argument that data from many educational settings are generated by random coëfficiënt processes. Therefore, we also believe that statistical inference should be based on the same kind of model." (1983, p. 107) .

De behoefte aan random coëfficiënt modellen wordt het duidelijkst beschreven in Aitkin en Longford (1986), die de verschillende variantie analytische modellen met elkaar vergelijken. Zij concluderen, dat deze modellen onmisbaar zijn in situaties waar het context effect "appreciable" is. (l.c. p. 26). Juist in genoemde situaties geven andere modellen misleidende informatie. Als voorbeeld geven zij de "serious overstatement" van de significantie van de contextuele variabele in een model, waarbij beide niveaus in de regressie vergelijking de variantie van de afhankelijke variabele verklaren. Deze "overstatement" is het gevolg van de onderschatting van de standaardfout in de regressie op het individuele niveau. In Kreft en Van den Eeden (1985) vinden we een soortgelijke opmerking naar aanleiding van het proefschrift van Van Marwijk Kooy.

Het lineaire model, random of fixed. Bij analyses van schooleffecten is het tot nu toe meest gebruikte model het lineaire model, en de meest gebruikte techniek multiple regressie analyse. Deze modellen worden in onderzoek van onderwijs ook wel aangeduid met "educational production function" of met de naam "input-output modellen". Het kenmerk van deze modellen is, dat eigenschappen van leerlingen en eigenschappen van scholen (de "inputs") gerelateerd worden aan de uitkomsten van het leerproces (de "outputs"). Een klassiek voorbeeld van een "single equation production function" is

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \epsilon_{ij} \quad (2)$$

Hierbij is y_{ij} de score van de schoolprestatie van leerling i op school j , x_{ij} is de

observatie van de onafhankelijke variabele op individueel of op schoolniveau van dezelfde leerling, α_j en β_j zijn de onbekende parameters, die geschat moeten worden, en ε_{ij} is de random errorterm, met de gebruikelijke veronderstellingen van een gemiddelde nul en een variantie σ^2 .

Ook in het random coëfficiënt model is het lineaire model het uitgangspunt van de analyse. Het bovenstaande model (2) kunnen we beschouwen als een speciaal geval van het algemene model, nl. een model, dat de schooleffecten als vaste coëfficiënten opneemt en geen correlatie veronderstelt van de individuele storingstermen binnen contexten. Dit model is daarom behoorlijk restrictief. Meer algemeen is het random coëfficiënt model, waarbij een grotere homogeniteit verondersteld wordt tussen populaties op eenzelfde school, dan binnen de gehele steekproef gesommeerd over alle scholen. Deze homogeniteit is in het model opgenomen in de context-storingsterm. Anders gezegd de covariantie van de storingstermen op het individuele niveau, binnen eenzelfde context, is een maat voor schooleffect, of voor homogeniteit, veroorzaakt door een gelijke onderwijssituatie. De vergelijking van het random coëfficiënt model op het individuele niveau, of zo men wilt als "single equation", is

$$y_{ij} = \alpha + \beta x_{ij} + u_j + \gamma_j x_{ij} + \varepsilon_{ij}. \quad (3)$$

Hierbij is α het intercept van de gemiddelde school (het vaste deel), en is u_j de bijbehorende storingsterm (het random deel). Beide termen behoren bij de stochastische variabele, die het intercept in deze modellen is. Hetzelfde geldt voor de helling, waarbij we een vast deel vinden in de parameter β , en een random deel in de storingsterm γ_j , die tevens gecorreleerd is met u_j . Beide contextuele storingstermen zijn ongecorreleerd met de storingsterm op het individuele niveau. Deze laatste heeft een variantie σ^2 . De gecorreleerdheid van de storingstermen binnen contexten wordt daardoor bepaald door zowel de variantie van de individuele storingsterm als door de gewogen contextuele variantie. Het is duidelijk dat zowel (2) als (3) speciale gevallen zijn van (1). Als we in

(1) aannemen, dat u_j en t_j nul zijn, dan hebben we het fixed effect model (2). Wanneer we aannemen in (1) dat de vaste delen van het intercept en de helling gelijk zijn, dus $\alpha_j = \alpha$ en $\beta_j = \beta$ voor alle scholen, dan hebben we model (3).

Als we het meest algemene model (1) als uitgangspunt nemen op het individuele niveau, dan betekent dit, dat de drie parameters van dit model, het intercept, de regressie coëfficiënt en de variantie van de individuele storingsterm "random" zijn. Dit betekent, dat niet gezocht wordt naar de "beste schatter" van intercept of regressiecoëfficiënt, maar naar de variatie van deze "constanten" over contexten, uitgedrukt in de variantie van de contextstoringstermen. De aannamen omtrent deze storingstermen zijn van belang, omdat in het lineaire model de eigenschappen van toets en schatters geheel bepaald wordt door de aannamen omtrent de storingstermen.

Slopes en intercepts "as outcomes". We geven in deze paragraaf een aantal basisbegrippen van het regressiemodel en de manier waarop we deze gebruiken in de beide stappen van onze analyse. De regressiecoëfficiënten van de binnen-scholen regressieanalyses uit de eerste stap worden in de tweede stap gebruikt als de afhankelijke variabelen in een regressieanalyse op schoolniveau. Deze modelmatige vorm van aggregatie vinden we al eerder in de literatuur. De procedure staat bekend onder de naam "slopes as outcomes", naar aanleiding van een artikel van Burstein, Linn en Capell, uit 1978. Om vergelijkbare redenen kan je ook spreken over "intercepts as outcomes". Dit laatste betekent, dat de afhankelijke variabele in de tweede stap het "overall" gemiddelde effect van de scholen uit de analyses van het individuele niveau zijn. Bij "slopes as outcomes" wordt het interactie effect tussen categorieën van leerlingen en de school als afhankelijke variabele gebruikt. Deze binnenklassen-regressie coëfficiënt is van belang bij evaluaties van beleid, dat erop gericht is achterstandssituaties van bepaalde categorieën leerlingen op te heffen. We denken hier met name aan de evaluatie van het voorrangbeleid. In de woorden van Burstein, Linn en Capell: "One way to take differences in within class slopes into account is to use slopes as outcomes in a between-class analysis" en verderop: "It should be noted that a linear regression with slopes as dependent variables is conceptually equivalent to a non-linear model in the

original dependent variable, that incorporates cross-product terms." (1978, p. 357).

In beide gevallen betekent dit, dat eerst per context een eenvoudige regressie berekend wordt. Er zijn n_j scores op de afhankelijke variabele op school j , en n_j scores op de onafhankelijke variabele (in ons voorbeeld resp. de individuele schoolloopbanen en de individuele CITO-scores). Voor iedere context apart geldt het volgende: het intercept (a_j) is het aanvangsniveau van de afhankelijke variabele. Het geeft de waarde aan bij $x=0$. Ongelijkheid in de intercepten tussen scholen betekent, dat alle leerlingen op de ene school hoger of lager scoren op de afhankelijke variabele. Het intercept is een algemene maat voor het schooleffect, en het wordt daarom in covariantie analyse gebruikt om verschillen tussen scholen te toetsen. De helling (b_j) geeft per school aan hoeveel een punt meer op de onafhankelijke variabele waard is op de afhankelijke variabele. Als de hellingen, vergeleken tussen contexten, significante verschillen vertonen, wijst dit op interactie tussen de context en individuele variabelen uit het model. Concreet betekent dit, dat leerlingen met eenzelfde score op een onafhankelijke variabele in de ene school beter uitkomen dan op een andere school.

Modellen, parameters en toetsen. Het voordeel van het random coëfficiënt model is duidelijk. In dit model worden de ongemeten variabelen van het individuele en het context niveau, die samengevat zijn in de storingstermen, uit elkaar gehaald. Deze modellen hebben echter nog een ander, niet onaanzienlijk voordeel boven vergelijkbare vaste effecten-modellen. Dit voordeel zit in de reductie van het aantal te schatten parameters. In het door ons in de volgende paragraaf geanalyseerde databestand van 106 scholen werden voor het ANCOVA model (fixed effect covariantie analyse) 108 parameters geschat. Dit grote aantal is nodig omdat voor alle 106 scholen een intercept geschat moet worden. Daarnaast hebben we nog een regressie coëfficiënt en de variantie van de storingsterm van het individuele niveau. Wanneer we regressiecoëfficiënten en varianties per school vrij laten, dan hebben we zelfs 318 parameters. Bij toepassing van het random coëfficiënt model (3) wordt het aantal te schatten parameters beperkt tot de drie varianties van de individuele storingsterm, van het intercept en van de regressiecoëfficiënt, tot de covariantie van de contextuele

storingstermen, en tot de twee fixed componenten van intercept en helling. Deze laatste modellen zijn daardoor aanzienlijk eenvoudiger. Er is een groot verschil tussen het schatten van 318 of 6 parameters.

Over het toetsen van modellen nog het volgende. De toetsen volgen min of meer vanzelf uit de maximum likelihood methoden, waarmee de parameters uit onze modellen worden geschat. Voor details verwijzen we naar De Leeuw en Kreft (1986). Elk model wordt getoetst ten opzichte van een nulmodel, of ten opzichte van een ander, minder restrictief model. Alleen het nulmodel zelf kan niet getoetst worden. Het nulmodel is het meest algemene en minst restrictieve model van de familie van modellen.

6. Modeltoetsen met voorbeelden.

Voor het fitten van de verschillende modellen gebruiken we gegevens uit het CBS databestand, het SMVO cohort 1977. We trekken hieruit een deelsteekproef van 3590 leerlingen verdeeld over 106 scholengemeenschappen voor mavo, havo en vwo. (Voor de wijze van steekproeftrekking en andere selectie procedures verwijzen we verder naar Kreft en De Leeuw, 1986). De door ons geselecteerde individuele achtergrondvariabele is de CITO-score op een verkorte versie van de reken- en taaltoets. De afhankelijke variabele is de speciaal voor dit doel geconstrueerde schoolloopbaanvariabele (zie ook hiervoor Kreft en De Leeuw, 1986). In onderstaande tabel staan de verschillende modellen met het aantal te schatten parameters. In deze tabel zien we ook de modeltoetsen, met bijbehorende χ^2 -waarden en vrijheidsgraden. In Tabel 1 verstaan we onder het nulmodel het model van formule (2), met $3 \times 106 = 318$ vrije parameters. Covariantieanalyse heeft $106 + 1 + 1 = 108$ vrije parameters, en als we veronderstellen dat de 106 intercepten gelijk zijn krijgen we het fixed model met gelijke intercepten, met nog $1 + 1 + 1 = 3$ parameters. In de contextuele modellen die in de tabel genoemd worden veronderstellen we dat de hellingen allemaal gelijk zijn, maar dat het intercept een lineaire functie is van een contextuele variabele, in ons geval de gemiddelde CITO score per school. Dus

$$\alpha_j = p + qx_j + u_j. \quad (4)$$

waarbij in de vaste effecten versie de storingsterm u_j verdwijnt. In formule (4) is x_j de gemiddelde CITO-score op school j . Het contextuele vaste model heeft daardoor 4 parameters (p , q , β , en σ^2), het random model heeft er 5 (ook nog ω^2 de variantie van de contextuele storingsterm u_j). Het gelijke intercepten random model tenslotte is dat uit formule (3) met t_j , de random component van de helling, gelijk aan nul. De vrije parameters zijn α , β , ω^2 , en σ^2 . In Tabel 2 vinden we schattingen van de vrije parameters.

Tabel 1: Samenvatting modeltoetsen.

MODEL	# VRIJE PARAMETERS	VERSCHIL LOG-LIKELIHOODS	CHI- KWADRAAT	DFR
(1) nulmodel fixed	318	****	****	****
(2) covariantie analyse	108	(1)-(2)	280.16	210
(3) gelijke intercepten fixed	3	(2)-(3)	437.40	105
(4) contextueel fixed	4	(2)-(4)	418.95	104
		(5)-(4)	162.58	1
(5) contextueel random	5	****	****	****
(6) gelijke intercepten random	4	(5)-(6)	4.45	1

Tabel 2:

Samenvatting modelschattingen.

nulmodel: intercepten, hellingen, en individuele varianties verschillen per school.covariantie analyse: intercepten verschillen per school.contextuele modellen: $p = -2.0820$ en $q = .0163$

MODEL	INTERCEPT	HELLING	VARIANTIESCHATTINGEN	
			INDIVIDUEEL	SCHOOL
nulmodel	****	****	****	.0
covariantie analyse	****	.0622	.5335	.0
gelijke intercepten fixed	-2.7410	.0599	.6026	.0
contextueel fixed	****	.0622	.5996	.0
contextueel random	****	.0622	.5496	1.6868
gelijke intercepten random	-2.8224	.0616	.5497	1.8012

De varianties en de covariantie van de beide individuele variabelen uitgesplitst naar tussen en binnen staan in Tabel 3. We kunnen in sommige gevallen zien hoe de parameterschattingen in Tabel 2 uit deze statistieken afgeleid zijn.

Tabel 3: Partitionering van de varianties en covarianties in tussen- en binnen-componenten.

	TOTAAL	TUSSEN	BINNEN	T/B
$v(x)$	95.1289	13.6509	81.4780	0.1435
$v(y)$	0.9436	0.0948	0.8488	0.1005
$c(x,y)$	5.6957	0.6265	5.0692	0.1100
β	0.0599	0.0459	0.0622	
σ^2	0.6026	2.2355	0.5334	

Onderaan Tabel 3 staan drie verschillende schattingen van regressie-coëfficiënten van de regressie van CITO-toets op schoolloopbaan, met de daarbij behorende schatting van de variantie van de individuele storingsterm. In iedere kolom is de beta berekend door de covariantie van xy te delen door de variantie van x . Dit levert de schatting op van het model, waarvan aangenomen wordt, dat alle intercepten en ook alle hellingen voor alle scholen gelijk zijn (het gelijke intercepten fixed model uit Tabel 1). In de tweede kolom van Tabel 3 zien we schattingen voor een ander veel gebruikt model in schooleffectiviteits onderzoek. In Nederland bijvoorbeeld in het proefschrift van Van Marwijk Kooy. In dit model is de "unit of analyses" de school, en de regressie is berekend over 106 scholen, waarbij de binnen scholen variantie wordt verwaarloosd. We zien dat de bij het model horende coëfficiënt onbetrouwbaar is, vanwege de grote foutenvariantie. In de derde kolom vinden we de schatter van de helling die hoort bij het covariantie analyse model.

De laatste twee modellen zijn modellen die schoolverschillen kunnen meten, maar niet verklaren (dat wil zeggen relateren aan andere variabelen die schoolverschillen beschrijven). Dit gebeurt wel in de contextuele modellen (Boyd en Iversen, 1979). Een

dergelijk model, te beschouwen als de tweede stap na de covariantie analyse, paste slecht op onze gegevens. Terwijl de daarvoor uitgevoerde covariantie analyse wel goed paste, zoals we kunnen zien aan de chi-kwadraat en vrijheidsgraden in Tabel 1 .

Een toepassing van een model onder de aanname van random coëfficiënten op dezelfde gegevens geeft ons de gelegenheid om binnen dit model te toetsen, of de aanname waaronder de vorige modellen getoetst zijn, juist is. We toetsen of ω^2 , de variantie van de context storingsterm, gelijk is aan nul. Dit levert een chi-kwadraat van 162.58 met een enkele vrijheidsgraad. Zie weer Tabel 1. De conclusie is hier gerechtvaardigd, dat voor onze gegevens, de storingstermen binnen contexten gecorreleerd zijn. Een traditioneel contextueel model past niet bij deze data.

Na het fitten van de verschillende modellen lijkt het erop dat we de variabiliteit van de intercepten kunnen 'verklaren' door binnen scholen gecorreleerde storingstermen aan te nemen. De correlatie tussen de storingstermen van leerlingen op eenzelfde school varieert van .05 (voor scholen met 60 leerlingen) tot .35 (voor scholen met 6 leerlingen). Anders gezegd: voor grote scholen kunnen we zeggen dat de variantie van de storingstermen 90% individueel 'bepaald' is, voor kleine scholen is wel 40% door de school 'bepaald'.

7. Referenties.

- Aitkin, M. A., & Longford, N.T. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the royal statistical society* (in press).
- Blalock, H.M. (1984). Contextual-effects models: Theoretical and methodological issues. *Annual review of sociology*, 10, 353-372.
- Boyd, L.H. & Iversen, G.R. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont, CA: Wadsworth.
- Burstein, L. (1978). Assessing differences between group and individual level regression coefficients: alternative approaches. *Sociological methods and research*, 7, 5-27.
- Burstein, L. , Linn, R.L., & Capell, F.J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of educational statistics*, 3, 347-383.
- Eeden, P van den, & Saris, W.E. (1984). Empirisch onderzoek naar multilevel uitpraken. *Mens en maatschappij*, 59, 165-178.
- Hannan, M.T. (1971). *Aggregation and disaggregation in sociology*. Lexington, MA: Heath-Lexington.
- Hauser, R.M. (1970). Context and consex: a cautionary tale. *American Journal of sociology*, 75, 645-663.
- Hauser, R.M. (1974), Contextual analysis revisited, *Sociological methods and research*, 2, 365-375.
- Harvard educational review*. Special issue, 51, 4, november, 1981.
- Hays, W.L. (1974) *Statistics for the social sciences*. Holt, Rinehart & Winston, London.
- Kreft, G.G.& Leeuw. J de. (1986). *Schooleffecten op schoolloopbanen*. Paper gepresenteerd op de gezamenlijke bijeenkomst van de SISWO-werkgroepen "Longitudinaal" en MULOLOG.
- Kreft, G.G. & Eeden, P. van den (1985). Valkuilen bij het proefschrift van L. van Marwijk Kooy, en hoe deze uit de weg te gaan. In : D.P. Kroes (red), *Studiebijeenkomst over de dissertatie " Scholen verschillen"*. SISWO-Publicatie 306. Amsterdam: SISWO.

- Leeuw, J. de. & Kreft, G.G. (1984). **Over multilevel analyse**. Leiden, The Netherlands: Department of Data Theory FSW/RUL.
- Leeuw, J. de. & Kreft, G.G. (1986). Random coefficient models for multilevel analysis. **Journal of educational statistics**, 11, 1, 57-85.
- Marwijk Kooy, L. (1984). **Scholen verschillen. Een verkennend onderzoek naar het intern functioneren van vijftwintig scholengemeenschappen vwo-havo-mavo**. Dissertatie Rotterdam.
- Pedhazur, E.J. (1975), Analytic methods in studies of educational effects. **Review of research in education**, 3, 243-286.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. **American sociological review**, 15, 351-357.
- Swanborn, P.G.(1973), **Variabelen en hun meting**. Boom, Meppel.
- Swanborn, P.G. (1978), Schoolkenmerken en leerlinggedrag. Enkele methodologische merkwaardigheden in het onderzoeksbedrijf. **Mens en maatschappij**, 53, 209-216.
- Tacq, J.J.A. (1986) **Van multiniveau probleem naar multiniveau analyse**. Paper gepresenteerd op de sociologendagen, april, 1986.
- Tate, R.L. & Wongbundit, Y. (1983). Random versus nonrandom coefficient models for multilevel analysis. **Journal of educational statistics**, 8, 103-120.