

Wat beïnvloedt het risico op voortijdig schoolverlaten?

Een multivariate analyse

Frank Pijpers

De meeste jongeren in Nederland hebben succes op school en behalen een startkwalificatie. Een kleine minderheid van de jongeren verlaat echter het onderwijs zonder een startkwalificatie, ofwel voortijdig. Er is steeds meer interesse in onderzoek naar de oorzaken van voortijdig schoolverlaten in het voortgezet onderwijs, middelbaar beroepsonderwijs, en het voortgezet algemeen volwassenen onderwijs. Univariate analyses, of analyses gebaseerd op steekproeven, kunnen een onvoldoende of vertekend beeld van de risicofactoren geven. Uit deze multivariate studie blijkt dat het voortijdig schoolverlaten beïnvloed wordt door een complex samenspel van factoren.

1. Inleiding

Nederland is in hoge mate een kenniseconomie en dat betekent dat het zeer belangrijk is voor jongeren om een startkwalificatie te hebben. Regeringsbeleid is er dan ook op gericht om (voortijdige) schooluitval te voorkómen. Als onderdeel hiervan is de kwalificatieplicht ingevoerd voor leerlingen tot 18 jaar. De doelstelling is om het aantal leerlingen van 22 jaar of jonger dat jaarlijks voortijdig de school verlaat, terug te dringen tot onder de 35 duizend. In de afgelopen jaren is het aandeel voortijdige schoolverlaters gestaag gedaald van 4,1 procent in 2005/06 naar 3,3 procent in 2008/09 (CBS, 2010). Gericht beleid dat effectief moet zijn in het terugdringen van voortijdig schoolverlaten heeft een goed beeld nodig van wat de grote risico factoren zijn die verband houden met voortijdig schoolverlaten.

Of een leerling de school voortijdig verlaat hangt meestal af van een complex samenspel van factoren. Univariate analyses geven hiervan vaak een onvoldoende beeld. Het niveau van kennis en vaardigheden waarmee het voortgezet onderwijs (vo) begonnen wordt, zoals dat gemeten kan worden door een Cito-toets score, zal ongetwijfeld invloed hebben op het succes. Het is echter mogelijk dat met dezelfde toetsscores de leeftijd van een leerling ook een rol speelt. Bij oudere leerlingen is namelijk een grotere tijd verstreken tussen het afnemen van een Cito-toets en deelname in het middelbaar beroeps onderwijs (mbo) of het voortgezet algemeen volwassenen onderwijs (vavo), waardoor voor hen een Cito-score wellicht een minder belangrijke voorspeller is voor schooluitval.

Het doel van dit artikel is om dit soort interacties in kaart te brengen. Het CBS heeft de beschikking over integrale gegevens uit registraties, waarop een multivariate analyse is gedaan. In het verleden is al wel multivariaat onderzoek gedaan maar daarbij is gebruik gemaakt van andere methoden en andere achtergrondkenmerken (Herweijer, 2008) of het werk was alleen gebaseerd op steekproeven (Meng et al, 2009).

De centrale vragen die een rol spelen in dit artikel zijn:

- Welke achtergrondkenmerken van leerlingen spelen een rol bij voortijdig schoolverlaten en hoe interacteren deze kenmerken?
- Wat is de invloed van scholen; bestaat er zoiets als meetbaar betere en slechtere instellingen voor wat betreft de schooluitval?

2. Methode

De analyse in dit artikel is gebaseerd op multi-niveau modellering (Engels: multilevel modelling of MLM). Het centrale idee van MLM is dat er een correlatie bestaat tussen een of meer afhankelijke variabelen en beïnvloedende factoren die eigenschappen zijn van individuele deelnemers. Echter, de deelnemers zijn in te delen in groepen, hier de onderwijsinstellingen, en tussen de verschillende instellingen kunnen er systematische verschuivingen zijn in de afhankelijke variabele: het risico van voortijdige schooluitval.

In het geval van voortijdig schoolverlaten zijn er diverse persoons- en huishoudenskenmerken die een invloed kunnen hebben op het risico op voortijdige schooluitval. Daarnaast kan ook de opleidingsinstelling een invloed hebben op het risico op voortijdig schoolverlaten. Omdat de leerlingpopulaties tussen deze instellingen juist verschillen op die persoons- en huishoudenskenmerken, moet daarvoor gecorrigeerd worden. Voor dit soort analyses is MLM ontworpen.

Het resultaat van de modellering is tweevoudig. In de eerste plaats geeft het inzicht in de effecten die de verschillende achtergrondkenmerken hebben op het risico van voortijdig schoolverlaten, en de interacties daartussen.

Verder geeft het model een 'voorspelling' voor de kans op al dan niet voortijdig schoolverlaten voor iedere deelnemer: een waarschijnlijkheid berekend op grond van de achtergrondkenmerken die in het model zijn opgenomen. Vanzelfsprekend is voor iedere deelnemer bekend of ze daadwerkelijk voortijdig de school hebben verlaten of juist niet. De voorspelling van het model kan vergeleken worden met de daadwerkelijke uitkomst. Voor groepen leerlingen van meer dan ongeveer 200 in aantal, liggen de voorspelde aandelen en de waargenomen aandelen heel dicht bij elkaar.

De populatie en de kenmerken

De analyse wordt uitgevoerd op een dataset met alle leerlingen van 12 tot 23 jaar in het vo, mbo, en vavo in het schooljaar 2007/08, die in de Gemeentelijke Basis Administratie zijn opgenomen. Dit laatste wordt gedaan om te voorkómen dat leerlingen die emigreren aangemerkt worden als vsv'er.

Het gaat dan om 1,33 miljoen deelnemers, waarvan er 48 duizend het onderwijs voortijdig hebben verlaten (3,7 procent).

Voor de deelnemers aan onderwijs zijn een aantal achtergrondkenmerken beschikbaar in de dataset van het CBS. Hieruit is een selectie gemaakt, waarvan wordt verwacht op grond van univariate analyses dat ze een dominante invloed hebben. De reden om een selectie te maken is dat het aantal coëfficiënten dat moet worden bepaald al heel snel oploopt, wanneer ook interacties (kruistermen) tussen achtergrondkenmerken in het model worden betrokken. In dit onderzoek zijn verder alleen paren van achtergrondkenmerken als interacties meegenomen, aangezien de interpretatie steeds complexer wordt bij het toevoegen van hogere orde interacties.

In het model zijn zeven persoonskenmerken en zeven kenmerken van het huishouden van leerlingen meegenomen, en ook eventuele interacties tussen deze termen. De persoonskenmerken zijn:

1. Geslacht
2. Herkomstgroepering
3. Generatie van herkomst
4. Leeftijd
5. Leeftijdsverschil met jaargroep
6. Cito-toets totaalscore
7. Verdacht zijn geweest van een misdrijf in de afgelopen 3 jaar

De huishoudenskenmerken zijn:

1. Aantal personen in huishouden
2. Soort huishouden (thuis- of uitwonend; een- of twe-oudergezin)
3. Aantal verdieners in huishouden
4. Inkomstenklasse (in groepen gerelateerd aan minimumloon)
5. Verandering in inkomstenklasse over de afgelopen 3 jaar
6. Aantal veranderingen in soort huishouden over de afgelopen 3 jaar
7. Aantal veranderingen in adres over de afgelopen 3 jaar

Al deze kenmerken worden geschaald, zodat de invloed ervan onderling vergeleken kan worden (zie paragraaf 5). Verder zijn er bij al deze factoren de volgende kanttekeningen over hun precieze gebruik in het model.

Het aandeel voortijdig schoolverlaters in iedere leeftijdsgroep verhoudt zich niet rechtsevenredig met de leeftijd, Daarom wordt niet alleen de geschaalde leeftijd zelf, maar ook een kwadratische en kubische term van leeftijd meegenomen in het model. Deze hogere orde termen zijn allebei afzonderlijk ook geschaald.

Het verschil in leeftijd tussen de individuele deelnemer en alle andere deelnemers in dezelfde onderwijsvorm in hetzelfde leerjaar of op hetzelfde niveau, heeft als landelijk gemiddelde een waarde die vrijwel gelijk is aan nul. Het belang van deze variabele is om te meten of de deelnemer in deze betekenis vertraagd (leerling is ouder) of versneld (leerling is jonger) de opleiding volgt, of een late instromer is, hetgeen een risico kan zijn voor voortijdig schoolverlaten. Behalve deze term zelf wordt ook een kwadratische term in het model opgenomen, om uit te kunnen drukken dat als

bijvoorbeeld leerlingen die vertraagd zijn een groter risico hebben om vsv'er te worden, dat nog niet automatisch inhoudt dat leerlingen die versneld hun opleiding doorlopen een evenredig kleiner risico lopen.

Voor ongeveer de helft van de populatie is er geen Cito-toetsresultaat. De toets is niet verplicht, en zelfs als de toets is afgenomen kan het zijn dat de gegevens niet beschikbaar zijn bij het CBS. Daarnaast heeft het CBS niet de Cito-scores van vóór 2003 beschikbaar. De ontbrekende waarden zijn geïmputeerd, uitgaand van andere leerlingen met hetzelfde geslacht en dezelfde onderwijsvorm, maar voor wie wél een Cito-score beschikbaar is. Na imputatie worden de Cito-scores genormaliseerd: het landelijk gemiddelde wordt afgetrokken, en er wordt gedeeld door de landelijke standaard deviatie. Dit heeft mede tot gevolg dat voor die deelnemers waarvoor de Cito-score moest worden ingevoegd door imputatie, de genormaliseerde score zometentoe identiek 0 dan toch klein is. Op deze manier wordt de invloed van de imputatie zo beperkt mogelijk gehouden. Behalve de genormaliseerde Cito-toetsscore zelf, is ook een genormaliseerde kwadratische term in het model opgenomen omdat het aandeel voortijdig schoolverlaters zich niet rechtsevenredig verhoudt met de Cito-score.

Gezamenlijk genomen geven deze achtergrondkenmerken, met eventuele hogere orde termen, 18 basistermen waarvoor het model een lineaire coëfficiënt bepaalt. Uit deze 18 basistermen kunnen 153 verschillende paren gevormd worden, maar sommige van deze paren zijn dan samengesteld uit een lagere en een hogere orde term van een en hetzelfde achtergrondkenmerk. Deze blijven buiten beschouwing, hetgeen betekent dat er 148 kruistermen overblijven. In totaal produceert dit meest eenvoudige model dus 166 coëfficiënten die al dan niet statistisch significant kunnen zijn.

3. Resultaten

Figuur 1 geeft een beeld van de coëfficiënten die uit het model komen. De termen op de diagonaal zijn de lineaire coëfficiënten van elke basisterm afzonderlijk. Een '++' in een cel geeft aan dat er een groot verhogend effect is op het risico voor voortijdig schoolverlaten. Zo heeft de basisterm 'aantal keren verdacht' een '++' op de diagonaal. Dat betekent dat het aantal keren dat een leerling verdacht werd van een misdrijf het risico op voortijdig schoolverlaten sterk omhoog duwt. Termen met een '+' hebben ook een verhogend effect, maar met een iets kleiner gewicht. Termen met een '-' of een '--' duwen het risico op voortijdig schoolverlaten juist omlaag. De kwadratische term van het leeftijdsverschil met de jaargroep heeft een '--' op de diagonaal en heeft op zichzelf staand een verlagend effect op het risico op voortijdige schooluitval. Echter in samenspel met de andere factoren heeft het uiteindelijk een verhogend effect. Termen met een '*' zijn klein in waarde (positief of negatief) maar zijn wel nog statistisch significant. Als een cel leeg is, is de coëfficiënt uit het model statistisch niet significant. Dat geldt bijvoorbeeld voor herkomstgroep en generatie als afzonderlijke basistermen.

1. MLM-coëfficiënten voor risico op voortijdig schoolverlaten

	Geslacht	Herkomst groep	Herkomst-generatie	Leeftijd	Leeftijd ²	Leeftijd ³	Verschil jaar-groep	Verschil jaar-groep ²	Cito-score	Cito-score ²	Aantal personen huishouden	Soort huishouden	Aantal verdieners huishouden	Keren verdacht geweest	Inkomensklasse	Verandering inkomen	Verandering huishouden	Verandering adres
Geslacht	*																	
Herkomstgroep	-																	
Herkomstgeneratie	*	-																
Leeftijd	-	+																
Leeftijd ²	*	*	*															
Leeftijd ³	*	*	*	X	-													
Verschil jaargroep		+	-	+	-													
Verschil jaargroep ²	-	++	-	++	++		X	-										
Cito-score	*	*	*	*	*	+			*									
Cito-score ²	*	*	*	+	+				X	-								
Aantal personen huishouden ³							*				*							
Soort huishouden	*	*		+	-			++	-	*		+						
Aantal verdieners huishouden				+	*													
Keren verdacht geweest			+				+	++	-	-	*			++				
Inkomensklasse	*		+	-		+	-	--	+		*			+				
Verandering inkomen			*	*	*	*	*	+						*				
Verandering huishouden			-	+				-						++				
Verandering adres			-	+	*	*		*			*						*	+

++ = groot en positief -- = groot en negatief X = buiten beschouwing
 + = minder groot en positief - = minder groot en negatief
 * = klein (positief of negatief) [leeg] = niet statistisch significant

N.B. Alleen de onderdriehoek is gevuld omdat de termen symmetrisch zijn. De cellen met een X blijven buiten beschouwing omdat het een interactie is van een lagere en hogere orde van een en hetzelfde achtergrondkenmerk.

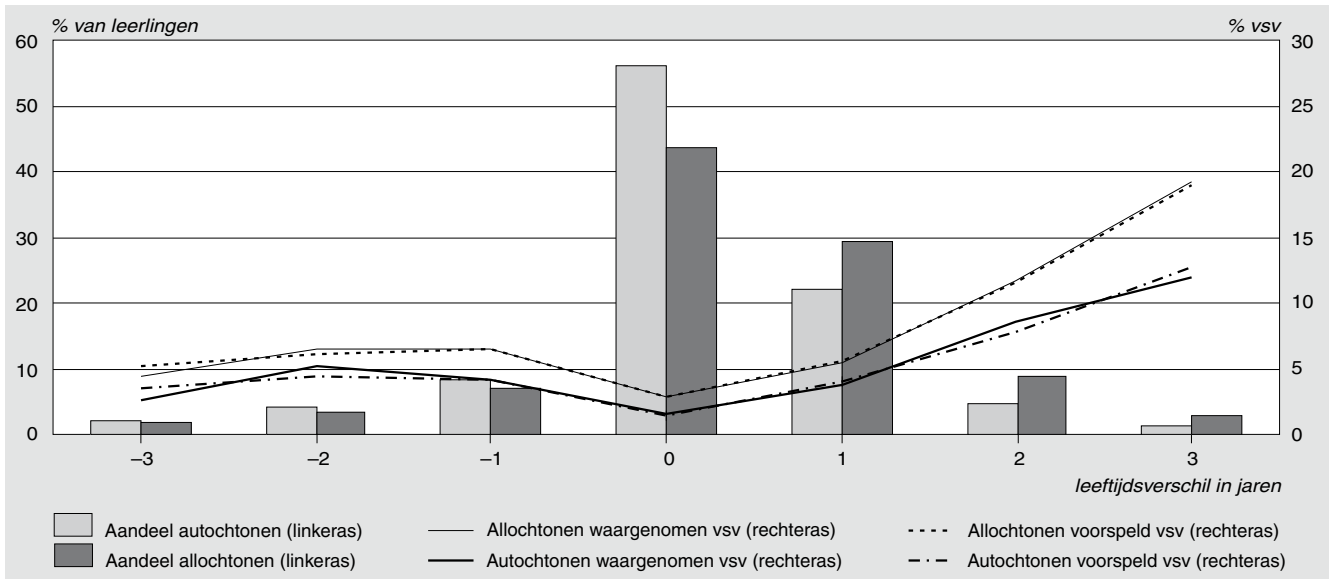
Vooraf interacties van kenmerken beïnvloeden risico op voortijdige schooluitval

Het geslacht als kenmerk op zich heeft maar een heel kleine invloed op het risico op voortijdig schoolverlaten. Herkomstgroep en generatie hebben op zichzelf helemaal geen effect. Het zijn juist de interacties tussen de achtergrondkenmerken die van belang zijn. Zo verlaten niet-westerse allochtonen weliswaar vaker het onderwijs zonder startkwalificatie, maar dit komt vooral door het gecombineerde effect van andere factoren, zoals leeftijd, het leeftijdsverschil met het gemiddelde van de klas, het aantal keren dat een leerling verdachte is geweest van een misdrijf, of het inkomstenniveau van het huishouden.

Leeftijdsverschil met jaargroep meer effect op voortijdig schoolverlaten bij allochtonen

Het leeftijdsverschil van een leerling met het gemiddelde van de groep leerlingen die in dezelfde onderwijsvorm en hetzelfde leerjaar zitten, verhoogt zowel bij autochtonen als bij allochtonen het risico op voortijdig schoolverlaten. Bij allochtonen is dat effect echter groter. Zo ligt het aandeel vsv bij allochtonen die een jaar ouder zijn dan het gemiddelde van de jaargroep op 5,5 procent. Bij de allochtonen die 3 jaar ouder zijn is dat zelfs bijna 20 procent. Bij autochtonen liggen die aandelen op respectievelijk op 3,8 en bijna 12 procent. Voor de leerlingen die jonger zijn dan gemiddeld is de relatie minder sterk.

2. Aandeel voortijdige schoolverlaters naar herkomst en leeftijdsverschil met jaargroep, met indicatie van de relative grootte van iedere groep

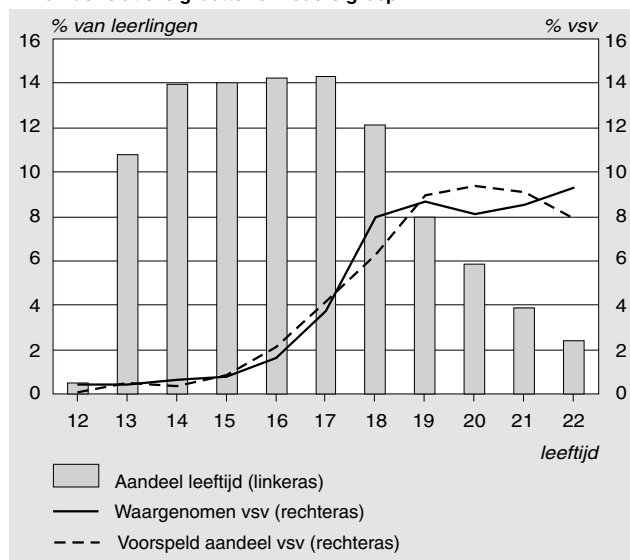


Figuur 2 toont zowel de waargenomen trend als de modelvoorspelling en deze liggen erg dicht bij elkaar. Dit is een specifiek voorbeeld dat met behulp van de achtergrondkenmerken over het algemeen zeer goed voorspeld kan worden wat het risico op voortijdige schooluitval van groepen leerlingen is.

Leeftijd van leerling speelt ook een rol

Naast het leeftijdsverschil met het gemiddelde van de klas, speelt de leeftijd zelf ook een grote rol. Dit is te verwachten, omdat de leeftijdsgrens van de leerplichtwet (16 jaar) een rol speelt, en verder jongeren in deze periode door hun puberteit heen gaan. Vanaf 16-jarige leeftijd loopt het risico op voortijdig schoolverlaten op. Zo ligt het aandeel dat voortijdig de school verlaat bij de 16-jarigen op ongeveer 2 procent. Bij de 19-jarigen is dat bijna 9 procent. Bij jongere leerlingen is het aandeel dat voortijdig de school verlaat

3. Aandeel voortijdige schoolverlaters naar leeftijd, met indicatie van de relatieve grootte van iedere groep



laag. Daarbij geldt ook dat jonge deelnemers die ook jonger zijn dan de gemiddelde leeftijd in hun klas, het effect op het risico op voortijdige schooluitval veel minder is.

Hoe vaker verdacht van een misdrijf hoe hoger het risico op voortijdige schooluitval

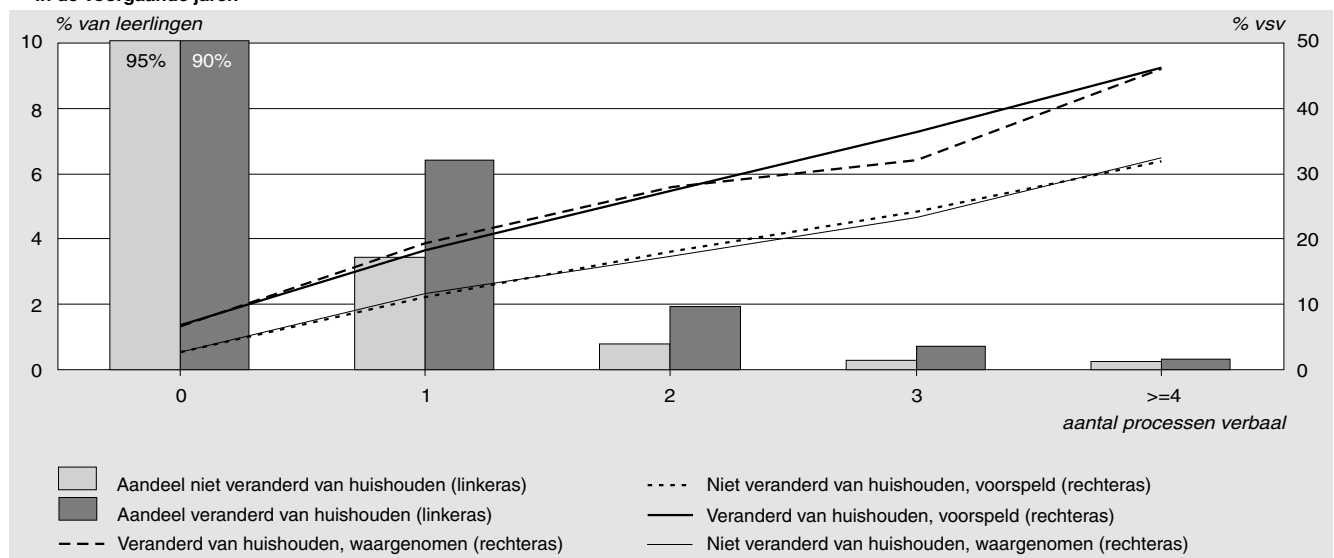
Daarnaast speelt het aantal keren dat een leerling gedurende het schooljaar en de twee jaren daarvoor verdachte is geweest van een misdrijf een grote rol. Hoe groter dit aantal is, hoe groter de kans op voortijdige schooluitval. Voor 95 procent van de leerlingen geldt echter dat zij nog nooit in aanraking zijn geweest met de politie.

De combinatie van veranderingen van huishouden en in aanraking zijn geweest met politie is een van de belangrijke kruistermen (zie figuur 1). Onder de leerlingen die een of meer keren van huishoudvorm zijn veranderd in de afgelopen 3 jaar is het aandeel dat een of meer keren een proces verbaal heeft gekregen groter. Het aandeel voortijdig schoolverlaters neemt al sterk toe met het aantal politiecontacten onder degenen die niet van huishoudvorm zijn veranderd, maar dat effect is nog sterker onder degenen die wel van huishoudvorm zijn veranderd.

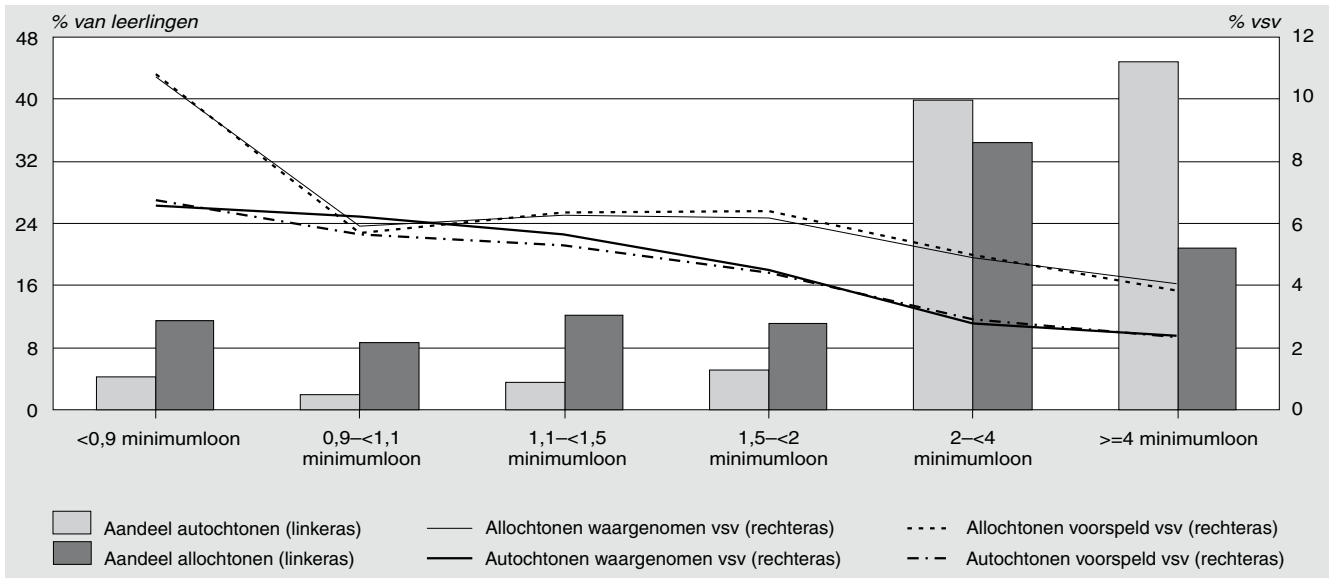
Laag inkomstenniveau verhoogt risico op voortijdige schooluitval meer bij allochtonen

Ook de inkomsten van het huishouden spelen een grote rol. Leerlingen uit een huishouden met een lager inkomstenniveau hebben een grotere kans op voortijdige schooluitval. Dit geldt zowel voor allochtonen als autochtonen. Bij inkomsten boven het minimumloon is echter de afname van het aandeel voortijdig schoolverlaters bij toenemend inkomstenniveau minder sterk onder allochtonen. Bovendien zijn de allochtonen ook meer vertegenwoordigd in de lagere inkomstenclassen. Binnen dezelfde inkomstenclasse is het verschil in aandeel voortijdige schooluitval tussen allochtone en autochtone leerlingen veel kleiner dan gemiddeld over alle inkomstenniveaus. Dit verschil dat overblijft heeft te maken met andere factoren: bijvoorbeeld een gemiddeld

4. Aandeel voortijdige schoolverlaters naar aantal veranderingen van huishouden en aantal keren verdacht zijn geweest van misdrijven in de voorgaande jaren



5. Aandeel voortijdige schoolverlaters naar inkomensklasse van huishouden en herkomst, met indicatie van de relatieve grootte van iedere groep



iets grotere vertraging die allochtonen hebben bij het doorlopen van hun schoolcarriere, of een taalbarriere die gemiddeld tot iets lagere Cito-scores kan leiden hetgeen vervolgens een verhoogde kans geeft op voortijdige schooluitval. Deze conclusie bevestigt eerdere bevindingen (Traag, 2007).

Verder speelt de Cito-score een rol. Een hogere score leidt tot een lagere schooluitval, maar dat effect geldt voornamelijk bij jongere leerlingen. De regelmaat waarmee leerlingen veranderen van vorm van huishouden of van adres over de jaren 2005, 2006, en 2007 voor leerlingen uit het schooljaar 2007/08 heeft ook een merkbaar gewicht. De overige factoren spelen wel een meetbare rol, maar die is ondergeschikt.

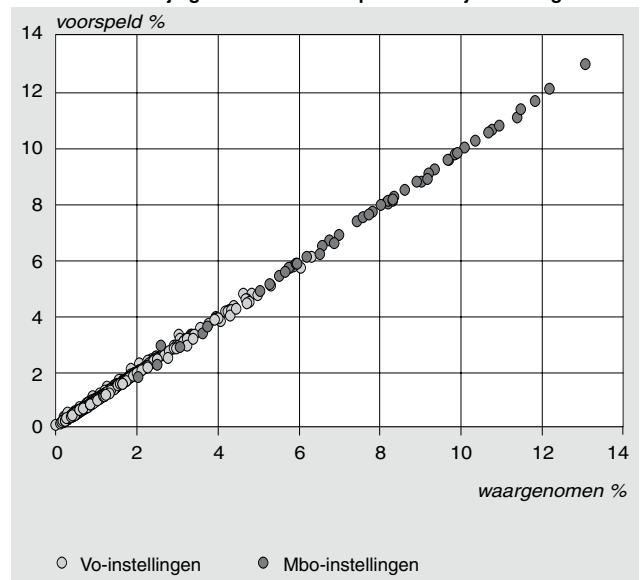
Effect van onderwijsinstelling lijkt gering

In het model kunnen de leerlingen gegroepeerd worden naar onderwijsinstelling. De scholen verschillen onderling in samenstelling van de leerlingpopulatie. De ene school heeft bijvoorbeeld meer allochtone leerlingen, terwijl een andere school juist meer vertraagde leerlingen heeft. Als de onderwijsinstellingen heel erg zouden verschillen in het effect dat ze hebben op het risico op voortijdige schooluitval, dan zouden er verschillen zijn tussen de waargenomen aandelen voortijdige schooluitval en de door het model voorspelde aandelen.

De voorspelde waarde van het model houdt namelijk rekening met de verschillen in achtergrondkenmerken die leerlingen hebben. Uit figuur 6 blijkt dat dit verschil tussen voorspeld en waargenomen vsv aandeel echter heel gering is. Als er bijvoorbeeld een paar scholen zouden zijn die helemaal niet effectief zijn in het voorkómen van vsv, zouden deze scholen in figuur 6 duidelijk onder de trendlijn verschijnen.

Het model geeft géén uitsluitsel over hoe effectief het beleid is, gemiddeld over alle instellingen. Dat kan alleen door ontwikkelingen van jaar tot jaar te beschouwen. Er kan wel

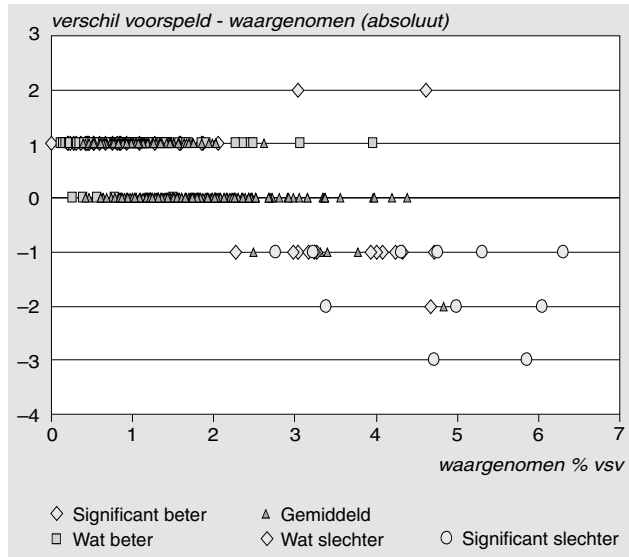
6. Aandeel voortijdige schoolverlaters per onderwijsinstelling



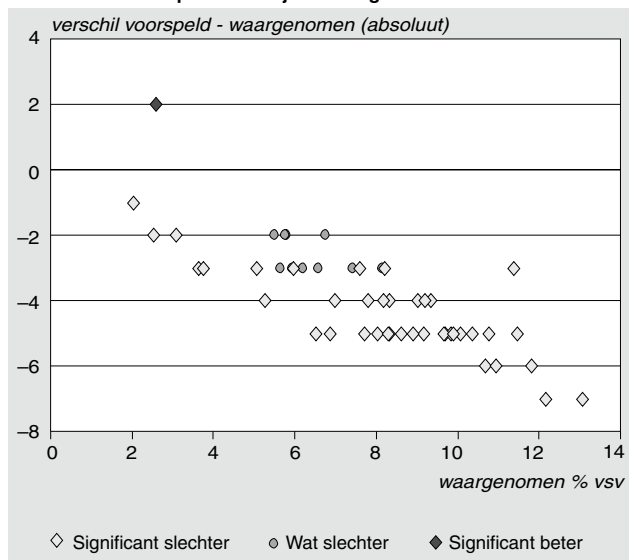
bekeken worden of bepaalde scholen veel beter of slechter zijn in het voorkómen van vsv, dan alle andere scholen. Uit de analyse is gebleken dat de daadwerkelijke verschillen in invloed op voortijdige schooluitval tussen instellingen beperkt zijn. Dit is echter wel meetbaar. Dit kan door per instelling te berekenen hoeveel leerlingen die voortijdig de school hebben verlaten het model meer of minder heeft voorspeld dan de daadwerkelijke aantallen. Het model geeft een schatting van hoeveel vsv-leerlingen een instelling zou moeten hebben, gezien de achtergrondkenmerken van de leerlingen die in het model zijn meegenomen. Het waargenomen aantal is wat de school daadwerkelijk aan vsv-leerlingen had in 2007/08. Het gaat hier dus om een vergelijking van absolute aantallen. Figuur 7 laat dit zien voor de instellingen in het voortgezet onderwijs en figuur 8 voor het mbo. In de figuren worden met verschillende symbolen de instellingen gepositioneerd die meetbaar ('significant') beter of slechter dan de voorspelling zijn, afgezet tegen hun aandeel voortijdige schooluitval van 2007/08. Het blijkt

echter in alle gevallen om kleine verschillen te gaan: in geen enkele instelling is het verschil meer dan 7 voortijdige schoolverlaters. Beleidsmatig zijn dit heel bescheiden aantallen. Van alle mbo-instellingen is er slechts één meetbaar ('significant') beter dan verwacht volgens de voorspelling. Alle overige mbo instellingen hebben meer vsv's dan verwacht zodat er geen mbo instellingen in de categorieën 'wat beter dan verwacht' of 'gemiddeld' vallen.

7. Verschil in waargenomen en voorspelde aantallen voortijdige schoolverlaters per onderwijsinstelling voortgezet onderwijs



8. Verschil in waargenomen en voorspelde aantallen voortijdige schoolverlaters per onderwijsinstelling Mbo



Alle onderwijsinstellingen hebben invloed op de kwaliteit van de begeleiding van (risico)leerlingen. Dalende vsv aandelen over de afgelopen jaren (CBS, 2010) hangen ongetwijfeld samen met de collectieve inspanningen van scholen om vsv te verlagen. Wat hier duidelijk wordt is dat er geen scholen zijn die veel effectiever, of veel minder effectief, zijn geweest dan anderen om de voortijdige

schooluitval verder terug te dringen. Dit stemt slechts gedeeltelijk overeen met het jaarrapport van de onderwijsinspectie (Inspectie van het Onderwijs, 2009) over het schooljaar 2007/08. Hierin wordt aangegeven dat de scholen in het voortgezet onderwijs met meer dan 3 procent voortijdige schooluitval vaak ook als zwak beoordeeld worden op andere criteria dan schooluitval. Uit figuur 7 blijkt dat bij ongeveer de helft van de scholen in het voortgezet onderwijs met een aandeel voortijdige schooluitval hoger dan 3 procent, dit resultaat inderdaad slechter is dan het model voorspelt. Met andere woorden zij hadden meer voortijdige schoolverlaters dan verwacht volgens de achtergrondkenmerken van de leerlingpopulatie. Voor de andere helft van deze groep scholen wijkt het uitvalspercentage echter niet af, of is het zelfs beter dan dat het model voorspelt op basis van de samenstelling van hun leerlingpopulatie. De grens van 3 procent selecteert scholen op grond van hun leerlingpopulatie in plaats van op hun effectiviteit in het voorkómen van schooluitval. Wel geeft ook de onderwijsinspectie aan dat nadere analyse van schooluitval noodzakelijk blijft.

4. Discussie

De multivariate analyse van voortijdig schoolverlaten met behulp van multilevel modellering geeft duidelijk aan dat vooral de combinaties van verschillende achtergrondkenmerken een rol spelen bij het risico op voortijdig schoolverlaten. Van de hier gebruikte kenmerken zijn, behalve de leeftijd zelf, de meest belangrijke factoren enerzijds het leeftijdsverschil tussen leerlingen en het gemiddelde van hun klasgenoten, en anderzijds het aantal keren in de voorgaande 3 jaren dat een leerling verdacht is geweest van een misdrijf. Interacties tussen deze drie factoren, en ook interacties met de overige modelfactoren blijken in zeer hoge mate bepalend te zijn voor het risico op voortijdige schooluitval. Zo is de relatief hogere uitval onder allochtone jongeren voor een groot deel een gevolg van andere factoren dan herkomst. Hier speelt bijvoorbeeld mee dat allochtone leerlingen beduidend vaker opgroeien in huishoudens met een lager inkomen, en wat vaker dan autochtonen ouder zijn dan de gemiddelde leeftijd van hun klasgenoten. De invloed van de individuele onderwijsinstelling op de voortijdige schooluitval blijkt gering te zijn. Er is veel eerder sprake van een breed maatschappelijke problematiek die een nationale aanpak blijft vereisen, in plaats van specifieke instellingen hierop aan te spreken.

Dat betekent echter niet dat alle verklaringen omtrent voortijdige schooluitval nu onderzocht zijn. Er zijn nog vele factoren van belang die vooralsnog ontbreken in het model. Bijvoorbeeld medische achtergrondkenmerken van leerlingen. Deelnemers kunnen door een chronische ziekte of andere zware aandoening, of door een ongeval of omdat ze slachtoffer zijn van een misdrijf, grote hindernissen hebben bij het vervolgen van hun opleiding en om deze reden tijdelijk of permanent uitvallen. Onderzoek wijst uit dat er ook correlaties bestaan tussen middelengebruik (alcohol of cannabis) en problemen op school, waarvan spijbelen en uitval uiteindelijke gevolgen kunnen zijn (ter Bogt et al., 2009). Naast deze voorbeelden van ontbrekende kenmerken van leerlingen, zijn er ook beperkingen van de

gegevens die wel beschikbaar zijn. Deelnemers die contacten hebben gehad met politie waarbij proces-verbaal is opgemaakt, kunnen niet worden onderscheiden naar de precieze aard van het misdrijf, noch of er daadwerkelijk justitiële gevolgen zijn geweest. Dit verlies van informatie heeft ongetwijfeld een nadelige invloed op de voorspellingskracht van deze factor.

In dit model is het voortgezet onderwijs samen met het mbo meegenomen. Aparte modellen voor beide onderwijssoorten zou verschillende uitkomsten kunnen opleveren.

Technische toelichting

Voortijdig schoolverlaters

Voortijdig schoolverlaters (vsv'ers) zijn in dit artikel leerlingen die op 1 oktober 2007 onderwijs volgen in het voortgezet onderwijs, het voortgezet algemeen volwassenen onderwijs, en het middelbaar beroepsonderwijs en die op 1 oktober 2008 het (bekostigd) onderwijs hebben verlaten en niet in het bezit zijn van een startkwalificatie; dat wil zeggen op zijn minst een havo- of vwo-diploma, of een diploma op mbo-2 niveau.

Het gaat hierbij om de voortijdig schoolverlaters met een leeftijd van 12 tot en met 22 jaar ultimo september 2008, en die op beide peilmomenten staan ingeschreven in de Gemeentelijke Basisadministratie. De uitkomsten hebben niet het doel het totaal aantal voortijdig schoolverlaters in Nederland weer te geven. Cijfers hierover zijn berekend door het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW, 2008).

Multivariate analyse methode

De implementatie van MLM geschiedt door het aanpassen van lineaire coëfficiënten voor de bepalende factoren $\{X_1, \dots, X_n\}$ en een hiërarchie van producten van deze factoren $\{X_1X_2, X_1X_3, \dots, X_2X_3, \dots, X_{n-1}X_n\}$. Gezamenlijk zijn deze coëfficiënten zodanig dat deze zo goed als mogelijk de afhankelijke variabele Y , de kans op vsv, reproduceren met een 'voorspelling' door een 'maximum likelihood' aanpak. Hierbij wordt verondersteld dat binnen iedere groepering van individuele onderwijsdeelnemers de statistische verdeling een normale verdeling is. Echter de verschillende groepen zullen een verschillende verwachtingswaarde $\langle Y \rangle$ en spreiding σ_Y hebben voor de afhankelijke variabele. Aangezien deze $\langle Y \rangle$ en σ_Y voor iedere groep tegelijkertijd bepaald worden met de lineaire coëfficiënten voor de factoren (en producten van factoren) is dit een niet-lineair minimaliserings probleem voor een χ^2 -maat over de populatie van onderwijsdeelnemers. Dit wordt opgelost door middel van een Levenberg-Marquardt routine, geïmplementeerd binnen SPSS.

Het resultaat van de modellering is een uitkomst voor het relatieve gewicht/belang van de verschillende bepalende kenmerken wanneer ze allemaal in samenspel beschouwd worden in plaats van ieder op zich. Verder geeft het model een 'voorspelling' voor de kans op al dan niet voortijdig schoolverlaten voor iedere deelnemer: een waarschijnlijk-

heid berekend op grond van de bepalende factoren die in het model zijn opgenomen. Het uitgangspunt hier is dat voor iedere deelnemer er een eindige kans is om vsv'er te worden, afhankelijk van heel veel factoren. Sommige daarvan kunnen achterhaald worden en opgenomen worden in een model, maar er blijven ook verborgen factoren. Het model kan niet gebruikt worden om voorspellingen te doen voor individuele leerlingen, maar alleen om het aandeel van vsv'ers over grote groepen leerlingen te voorspellen.

Het meest complete gegevensbestand inclusief achtergrondkenmerken dat beschikbaar is ten tijde van dit onderzoek betreft leerlingen die in schooljaar 2007/08 in onderwijs waren, waarvan een gedeelte in schooljaar 2008/09 zonder startkwalificatie was en niet deelnam aan het onderwijs, en daarmee dus voortijdig schoolverlater was. Het model is op dit schooljaar gebaseerd. De resultaten zullen echter nagenoeg onafhankelijk zijn van het schooljaar.

Schalen van modelfactoren

Om het relatieve belang van de diverse achtergrondkenmerken te kunnen aflezen in de resultaten van het model, is het nodig dat de individuele kenmerken zoveel als mogelijk een vergelijkbare orde van grootte hebben. Een van de kenmerken kan bijvoorbeeld waarden aannemen van 1 tot 10 duizend, en een ander kenmerk slechts waarden van 0 tot 0,01. Hiermee moet rekening worden gehouden als de waarde van de betreffende lineaire coëfficiënten uit het model met elkaar worden vergeleken. De manier waarop dit bereikt wordt, verschilt per kenmerk: voor sommige wordt een normaliserende schaling gebruikt, voor de overigen een bereikschaling.

1. Een normaliserende schaling waarbij voor iedere deelnemer het landelijk gemiddelde van de betreffende variabele wordt afgetrokken van de waarde die voor die deelnemer geldt, en dit verschil wordt vervolgens gedeeld door de (landelijke) standaarddeviatie van die variabele. Dit is zinvol voor variabelen die continu of fijnschalig zijn en/of waarvan bekend is dat de verdeling van de waarden bij benadering normaal is. De leeftijd van leerlingen is een voorbeeld hiervan.
2. Een schaling van het volle domein van de variabele in kwestie, zodat de geschaalde versie minimaal 0, en maximaal 1 is. Dit is zinvol voor variabelen die niet bij benadering normaal verdeeld zijn, of die bijvoorbeeld alleen maar een numerieke codering zijn van specifieke groeperingen van onderwijs deelnemers, zoals herkomst.

Onderwijsinstellingen

Bij de vergelijking van de modelvoorspellingen met de waargenomen cijfers over voortijdig schoolverlaten per onderwijsinstelling, zijn instellingen met minder dan 200 leerlingen in 2007/08 uitgesloten. De reden om kleine scholen uit te sluiten is dat de voorspellingen statistisch van aard zijn. Wanneer de groepen te klein worden is de voorspelling niet langer betrouwbaar. Per instelling zijn de VAVO-leerlingen altijd zodanig kleine aantallen dat voor deze groep de voorspellingen veel minder betrouwbaar zijn, dus deze zijn als gehele groep uitgesloten van de analyse.

Literatuur

- CBS (2010). *Jaarboek onderwijs in cijfers 2010*. Den Haag/Heerlen: Centraal Bureau voor de Statistiek.
- Herweijer, L. (2008). *Gestruikeld voor de start, de school verlaten zonder startkwalificatie*. Den Haag: Sociaal en Cultureel Planbureau.
- OCW (2008). *VSV-atlas totaaloverzicht Nederland. Resultaat convenantjaar 2007–2008*, 2^e editie.
- Traag, T., Velden, R.K.W. van der (2007). Voortijdig schoolverlaten in het VMBO. De rol van individuele kenmerken, gezinsfactoren en de school. *Sociaaleconomische trends*, 2/2007, 16–21.
- Inspectie van het Onderwijs (2009). *De staat van het onderwijs. Onderwijsverslag 2007/2008*. Utrecht: Inspectie van het Onderwijs.
- Meng, C., Coenen, J., Ramaekers, G., & Büchner, C. (2009). *Zonder diploma: Aanleiding, Kansen en Toekomstintenties*. Maastricht: Researchcentrum voor Onderwijs en Arbeidsmarkt.
- Bogt, T. ter, Lieshout, M. van, Doornwaard, S., Eijkmans, Y. (2009). *Middelengebruik en voortijdig schoolverlaten*. Universiteit Utrecht en Trimbos instituut.