



Panteia

Research to Progress

Research voor Beleid | EIM | NEA | IOO | Stratus | IPM



Kans op Stage 2017

Onderzoeksverantwoording

Jan de Kok

Zoetermeer, 13 maart 2017

De verantwoordelijkheid voor de inhoud berust bij Panteia. Het gebruik van cijfers en/of teksten als toelichting of ondersteuning in artikelen, scripties en boeken is toegestaan mits de bron duidelijk wordt vermeld. Vermenigvuldigen en/of openbaarmaking in welke vorm ook, alsmede opslag in een retrieval system, is uitsluitend toegestaan na schriftelijke toestemming van Panteia. Panteia aanvaardt geen aansprakelijkheid voor drukfouten en/of andere onvolkomenheden.

The responsibility for the contents of this report lies with Panteia. Quoting numbers or text in papers, essays and books is permitted only when the source is clearly mentioned. No part of this publication may be copied and/or published in any form or by any means, or stored in a retrieval system, without the prior written permission of Panteia. Panteia does not accept responsibility for printing errors and/or other imperfections.

Inhoudsopgave

1	Inleiding	5
1.1	Aanleiding	5
1.2	Opzet van deze onderzoeksverantwoording	5
2	Uitgangspunten: het onderliggende model	7
2.1	Het model	7
2.2	De kans op stage voor 2017	8
2.3	Kans op stage in zeven stappen	9
2.4	Kans op stage voor alle opleidingen bepaald	13
3	De gebruikte data	15
3.1	Registratiegegevens	15
3.2	Enquête onder erkende leerbedrijven	17
4	Complicaties en correctiefactoren	21
4.1	Correctiefactor ontbreken leerweg	21
4.2	Correctiefactoren inschrijvingen en ontbreken leerbedrijven	21
4.3	Correctiefactor uren_week	22
4.4	Correctiefactor omvang BPVOs	22
4.5	Correctiefactor dossierniveau	22
4.6	Geen correctie voor toe- en uittrading	23
5	De methodiek verder uitgewerkt	25
5.1	_1_hernoem.sps	25
5.2	_2_brongegevens vacatures corrigeren.sps	26
5.3	_3_weegfactoren leerbedrijvenquête.sps	28
5.4	_4_leerbedrijvenquête.sps	29
5.5	_5_correctiefactor dossier.sps	33
5.6	_6_aantal_inschrijvingen.sps	34
5.7	_7_omvang BPVOs.sps	35
5.8	_8_aanbod_jaar_Tplus1.sps	35
5.9	_9_vraag_jaar_T.sps	36
5.10	_10_vraag_jaar_Tplus1.sps	37
5.11	_11_regionale mobiliteit.sps	37
5.12	_12_indicatoren.sps	38



1 Inleiding

1.1 Aanleiding

SBB brengt jaarlijks de kans op stage in beeld voor de belangrijkste MBO opleidingen in Nederland. Door deze kans per kwalificatie (crebocode), type leerweg (BOL of BBL) en regio (35 arbeidsmarktregio's) te bepalen ontstaat een zeer gedetailleerd beeld van de kansen van MBO studenten om in het komende jaar een stageplaats te vinden.

In het verleden is de kans op stage bepaald op basis van de meningen en inzichten van de adviseurs en sectorspecialisten van SBB. Beschikbare informatie over aantallen studenten (per opleiding, type leerweg en/of regio) werd hierbij niet op een systematische en uniforme manier gebruikt.

In 2016 heeft SBB besloten om de kans op stage op een kwantitatieve manier te bepalen, met beschikbare registratiedata als uitgangspunt. Omdat niet alle benodigde informatie in deze registraties beschikbaar is, is een enquête onder erkende leerbedrijven gebruikt om additionele informatie te verzamelen. Deze werkwijze heeft een aantal voordelen ten opzichte van de manier waarop de kans op stage in het verleden is bepaald:

- De werkwijze is uniform: de kans op stage wordt voor alle opleidingen, alle typen leerwegen en alle arbeidsmarktregio's op dezelfde manier bepaald. De resultaten zijn hierdoor goed vergelijkbaar voor alle opleidingen.
- De werkwijze is transparant: door uitleg te geven over hoe de kans op stage berekend wordt (onder andere via deze onderzoeksverantwoording) kan iedereen nagaan hoe dit gedaan is. Dit vergemakkelijkt (interne en externe) discussies over de gevolgde werkwijze, waardoor in de toekomst verbeteringen eerder gerealiseerd kunnen worden.
- De werkwijze is geautomatiseerd: de jaarlijkse updates van de kans op stage kunnen hierdoor snel gerealiseerd worden.

2017 is het eerste jaar waarvoor deze nieuwe werkwijze is ontwikkeld. De huidige werkwijze kan gezien worden als een basisversie die het komende jaar verder ontwikkeld kan worden.

1.2 Opzet van deze onderzoeksverantwoording

In het voorjaar van 2016 heeft SBB de hoofdlijnen bepaald van de manier waarop de kans op stage bepaald zou moeten worden: wat moet de kans op stage weergeven, wat zijn de onderliggende indicatoren die de kans op stage bepalen, en welke informatie is nodig om deze indicatoren te kunnen bepalen.

De kern van de methodiek is simpel: bepaal (per opleiding, type leerweg en arbeidsmarktregio) het verwachte aanbod en de verwachte vraag naar stages en leerbanen, en vergelijk die met elkaar. Er zijn echter een aantal complicerende factoren waardoor het lastig is om beschikbare data over vraag en aanbod direct met elkaar te vergelijken. In hoofdstuk 2 worden de hoofdlijnen van de methodiek besproken en worden de belangrijkste complicerende factoren vermeld.

Veel van deze complicaties hangen samen met de kenmerken van de beschikbare gegevens waarop de kans op stage berekend kan worden. Daarom worden in hoofdstuk drie de beschikbare databronnen besproken, voordat in hoofdstuk vier verder ingegaan wordt op de complicerende factoren (en hoe hiervoor wordt gecorrigeerd).



De methodiek bestaat uit een aantal SPSS syntaxen. Deze zijn zo opgezet dat ze met een minimum aan aanpassingen elk jaar opnieuw gebruikt kunnen worden. In hoofdstuk 5 bespreken we de opzet van elk van deze syntaxen. De syntaxen zelf zijn aan SBB overgedragen.

Leeswijzer

Om een overzicht te krijgen van de gebruikte methodiek is het voldoende om hoofdstuk 2 te lezen. Wie inzicht wil krijgen in de verschillende complicaties en de oplossingen die daarvoor gekozen zijn, kan daarna hoofdstukken 3 en 4 lezen. Hoofdstuk 5 is alleen relevant voor de lezers die zich willen verdiepen in hoe de methodiek (inclusief de correcties) is uitgewerkt en geprogrammeerd.



2 Uitgangspunten: het onderliggende model

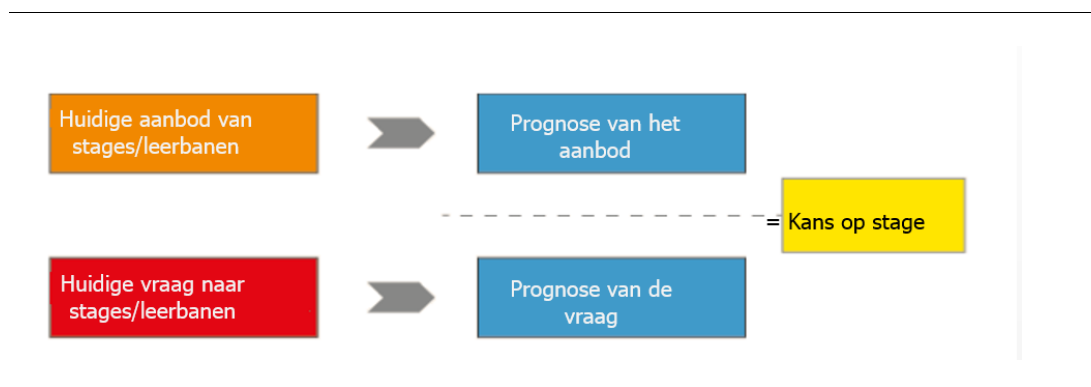
2.1 Het model

De kans op stage is door SBB als volgt gedefinieerd:

“Kans op stage is de graadmeter die een uitspraak doet over de verwachte kansen van mbo-studenten op het vinden van een stage of leerbaan. De graadmeter geeft de verhouding weer tussen het aantal stagezoekende studenten op school en het aanbod van stages en leerbanen bij bedrijven”
(Bron: SBB)

Om de kans op stage te bepalen, moet een prognose gemaakt worden van zowel het aanbod van stages en leerbanen als de vraag naar stages en leerbanen. Het aanbod komt van door SBB erkende leerbedrijven, de vraag komt van studenten die ingeschreven zijn bij publiek bekostigde MBO instellingen. De kern van het model wordt in figuur 1 weergegeven.

figuur 1 Kans op stage: de werkwijze op hoofdlijnen



Bron: SBB, 2016

2.1.1 Vier indicatoren voor de kans op stage

Dit is de eerste keer dat de kans op stage op basis van (vooral) registratiedata wordt bepaald. Het was vooraf niet duidelijk hoe de resultaten uit zouden pakken: hoe zouden de resultaten volgens deze nieuwe methodiek zich verhouden tot de oude methodiek, en hoe herkenbaar zouden de resultaten van de nieuwe methodiek voor de adviseurs zijn. Daarom is besloten om niet één, maar vier verschillende indicatoren af te leiden die allemaal inzicht kunnen geven in de (regionale spreiding van de) kans op stage (zie figuur 2).

De eerste twee indicatoren zijn gebaseerd op het model uit figuur 1. Het verschil tussen beide indicatoren is de manier waarop de prognose van het aanbod wordt bepaald (meer hierover in paragraaf 2.3.2).

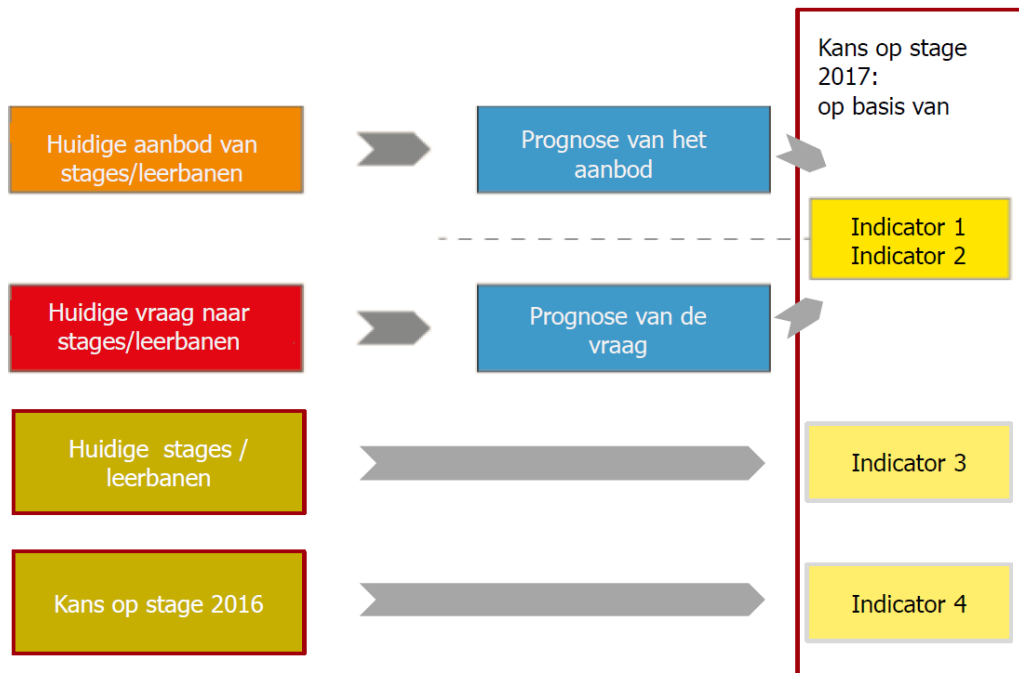
Deze twee indicatoren kunnen per arbeidsmarktregio bepaald worden (en daarmee verschillen tussen arbeidsmarktregio's weergeven), maar met één belangrijke maatstaf voor verschillen tussen regio's houden ze geen rekening: de regionale mobiliteit. Als derde indicator wordt daarom de regionale mobiliteit berekend, door per regio (en kwalificatie en type leerweg) te kijken naar de verhouding tussen het aantal



studenten dat in die regio woont en stage loopt bij een bedrijf buiten die regio, en het aantal studenten dat in een andere regio woont en stage loopt bij een bedrijf in die regio. Deze indicator kan bepaald worden op basis van informatie over alle stages en leerbanen die er in het afgelopen schooljaar zijn geweest.

Tot slot is het belangrijk om de resultaten van deze nieuwe methodiek te vergelijken met de resultaten van vorig jaar. Als vierde indicator wordt daarom de kans op stage gebruikt zoals die vorig jaar is bepaald.

figuur 2 Kans op stage: de werkwijze op hoofdlijnen



Bron: SBB, 2016

De definitieve, door SBB te publiceren kans op stage kan vervolgens berekend worden als een gewogen gemiddelde van deze vier indicatoren. De keuze voor de gewichten is dit jaar op een pragmatische manier bepaald, op basis van hoe betrouwbaar en realistisch de resultaten van de vier verschillende indicatoren ingeschat worden door enkele experts van SBB. Vervolgens is, ter verificatie, aan de adviseurs van SBB gevraagd hoe herkenbaar zij de aldus berekende kans op stage vinden.

2.2 De kans op stage voor 2017

De kans op stage wordt aan het begin van elk kalenderjaar gepubliceerd. De prognose betreft daarom de prognose voor het komende kalenderjaar; ten tijde van het schrijven van dit rapport was dat 2017. De huidige vraag en aanbod verwijzen in dit geval naar vraag en aanbod in 2016.

Veel van de databronnen waar de kans op stage op gebaseerd wordt (zie volgend hoofdstuk) zijn echter ingedeeld in schooljaren in plaats van kalenderjaren. Berekeningen op basis van deze data hebben daarom in eerste instantie betrekking op



schooljaren, niet kalenderjaren. Het gevolg hiervan is dat de berekende prognoses betrekking hebben op het komende schooljaar (in dit geval schooljaar 2016 / 2017). We beschouwen dit als de beste prognose voor het kalenderjaar 2017.

Een alternatieve oplossing zou zijn om niet alleen prognoses voor 2016 / 2017 te maken, maar ook voor 2017 / 2018. De kans op stage kan dan in eerste instantie voor beide schooljaren berekend worden. De kans op stage voor het kalenderjaar 2017 kan vervolgens berekend worden als het gemiddelde van de 'kansen' voor de twee opeenvolgende schooljaren. Theoretisch gezien is dit wellicht zuiverder, maar de resultaten worden minder precies omdat er extra bronnen van onzekerheid worden ingebouwd (omdat prognose voor 2 jaar gemaakt moeten worden in plaats van voor 1 jaar). Daarom hebben we niet voor deze alternatieve oplossing gekozen.

2.3 Kans op stage in zeven stappen

De kans op stage wordt in zeven stappen bepaald. Deze stappen worden nu kort besproken.

2.3.1 Bepaal het huidige aanbod van stages en leerbanen

Het aanbod van stages en leerbanen op een bepaald moment valt in twee delen uiteen: het aanbod dat op dat moment is vervuld en het aanbod dat op dat moment (nog) niet is vervuld. Het aanbod dat al vervuld is, bestaat uit stages en leerbanen waarvoor een BPVO (beroepspraktijkvormingsovereenkomst) is afgesloten. Daarnaast kunnen erkende leerbedrijven vacatures hebben openstaan die nog niet zijn ingevuld.

Bij de aanvang van dit project hadden de meest recente gegevens over het aanbod van stages en leerbanen betrekking op het begin van het lopende schooljaar (2016 / 2017). Het huidige aanbod kan bepaald worden op basis van de volgende informatie:

- Uit registraties over alle BPVOs van het afgelopen schooljaar (2015 / 2016) kan afgeleid worden hoeveel BPVOs aan het einde van dat schooljaar (en dus aan het begin van schooljaar 2016 / 2017) nog lopen.
- Erkende leerbedrijven kunnen hun vacatures in Artus registreren (een online registratiesysteem van SBB). Uit Artus kan het aantal openstaande vacatures aan het begin van het schooljaar 2016 / 2017 gehaald worden.

2.3.2 Maak een prognose van het aanbod van stages en leerbanen

Indicator 1:

Het huidige aanbod is een standcijfer: het totaal aantal stages en leerbanen dat aan het begin van het schooljaar 2016 / 2017 wordt aangeboden. In de loop van dit schooljaar zal dit aanbod uitgebreid worden. Als tweede stap moet daarom een inschatting gemaakt worden van het totale aanbod voor het schooljaar 2016 / 2017: het totale aantal stages en leerbanen dat in de loop van dit schooljaar door erkende leerbedrijven aangeboden zal worden. Per kwalificatie, type leerweg en arbeidsmarktregio moet bepaald worden in welke mate het huidige aanbod zal toenemen. Deze toename kan van twee groepen bedrijven komen:

- bedrijven die aan het begin van het schooljaar al een aanbod van stages en/of leerbanen hadden (in de vorm van lopende BPVOs en/of in Artus geregistreerde vacatures): zij kunnen in de loop van het schooljaar nog meer BPVOs afsluiten en/of vacatures openen;
- bedrijven die aan het begin van het schooljaar nog geen aanbod van stages en/of leerbanen hadden: zij kunnen in de loop van het schooljaar alsnog stages en/of leerbanen aanbieden (in de vorm van getekende BPVOs en/of openstaande vacatures).



Voor elk van deze groepen wordt een aparte correctiefactor berekend die de verhouding tussen het huidige aanbod en het verwachte aanbod weergeeft: voor de eerste groep is dit de correctiefactor inschrijvingen, voor de tweede groep de correctiefactor ontbreken leerbedrijven (zie paragrafen 4.1 en 5.4).

Indicator 2:

Vooraf waren er twijfels over de betrouwbaarheid van indicator 1. Daarom is een alternatieve maatstaf bedacht: ga niet uit van het aantal vacatures en lopende BPVOs aan het begin van het schooljaar, maar van het 'evenwicht' in het voorafgaande schooljaar (het aantal afgesloten BPVOs). Via een enquête onder leerbedrijven kan bepaald worden in welke mate de leerbedrijven verwachten dat het aantal BPVOs zal gaan toenemen. Deze "groeivoet BPVOs" kan gebruikt worden om een alternatieve prognose van het aanbod van stages en leerbanen te maken (in termen van aantal BPVOs).

2.3.3 Bepaal de huidige vraag naar stages en leerbanen

De vraag naar stages en leerbanen wordt uitgeoefend door studenten die ingeschreven staan bij een publiek bekostigde MBO opleiding in Nederland. De meest recente gegevens over aantallen ingeschreven studenten hebben betrekking op het voorafgaande schooljaar (2015 / 2016).

Niet alle studenten hoeven elk jaar een stage of leerbaan te volgen. Het doel van deze stap is daarom om te bepalen welk deel van de ingeschreven studenten in het voorafgaande schooljaar daadwerkelijk een stage of leerbaan heeft gelopen: wat is in het schooljaar 2015 / 2016 de verhouding tussen het aantal studenten en het aantal afgesloten BPVOs (per kwalificatie, type leerweg en regio).

2.3.4 Maak een prognose van de vraag naar stages en leerbanen

Om een prognose te maken voor de vraag naar stages en leerbanen in het schooljaar 2016 / 2017, moeten we per kwalificatie, type leerweg en arbeidsmarktregio een inschatting maken van de groei in het aantal afgesloten BPVOs.

Dit kan op verschillende manieren. Informatie over het aantal ingeschreven studenten en het aantal afgesloten BPVOs is voor meerdere jaren bekend. Een mogelijkheid zou zijn om een trend te schatten (bijvoorbeeld via regressieanalyses) over de ontwikkeling van het aantal afgesloten BPVOs en op basis hiervan een prognose te maken voor 2017. Dit kan per kwalificatie, type leerweg en/of regio gedaan worden (maar niet naar alle drie de dimensies tegelijkertijd). Een nadeel van deze aanpak is dat het complex (en daarmee tijdrovend) is en het aantal jaren waarover informatie beschikbaar is beperkt is.

Een andere manier is om de prognose te baseren op basis van demografische ontwikkelingen. Immers: veranderingen in het aantal afgesloten BPVOs kunnen door twee ontwikkelingen ontstaan:

- veranderingen in het aantal afgesloten BPVOs per leerling (per leerjaar);
- verandering in aantal studenten (per leerjaar).

Voor de kans op stage kijken we maar 1 jaar vooruit. Het lijkt redelijk om aan te nemen dat het aantal afgesloten BPVOs per leerling (per leerjaar) niet verandert. Veranderingen in de vraag naar stages en leerbanen worden dan bepaald door veranderingen in het aantal studenten (per leerjaar).



Het ministerie van OCW publiceert jaarlijks ramingen over het aantal studenten in het MBO, naar type leerweg en sector (zie hoofdstuk 3). Deze ramingen zijn gebruikt om een prognose te maken van de vraag naar stages en leerbanen voor 2017.

Ruimte voor verbetering

De prognose van de vraag naar stages en leerbanen is gebaseerd op de aanname dat het aantal afgesloten BPVOs per leerling (per leerjaar) niet verandert. Deze aanname zou bij een volgende editie van de Kans op Stage onderzocht kunnen worden.

De ramingen van het ministerie van OCW die gebruikt zijn, zijn alleen op hoog aggregatieniveau beschikbaar. Hierdoor wijken de prognoses nauwelijks af van de huidige vraag naar stages en leerbanen. Bij een volgende editie van de Kans op Stage zou onderzocht kunnen worden of er andere bronnen van informatie beschikbaar zijn, waarna alternatieve manieren om de prognose voor de vraag af te leiden onderzocht zouden kunnen worden.

2.3.5 Bereken de eerste twee indicatoren

De eerste twee indicatoren worden berekend door per kwalificatie, type leerweg en arbeidsmarktregio de prognose van het aanbod te delen door de prognose van de vraag. Hiervoor is het noodzakelijk dat vraag en aanbod op een vergelijkbare manier gedefinieerd worden. Het is niet voldoende om enkel naar aantallen BPVOs en/of vacatures te kijken, want er is een groot verschil in de omvang van BPVOs (zowel tussen verschillende kwalificaties als tussen BOL en BBL). Daarom moeten zowel de vraag als het aanbod in uren worden uitgedrukt. Voor de vraag is daar voldoende informatie voor beschikbaar, voor een deel van het aanbod ontbreekt informatie hierover in de beschikbare data. Hiervoor moeten aanvullende gegevens verzameld worden.

Deze twee indicatoren worden zo berekend dat ze aan de volgende randvoorwaarden voldoen:

- Ze zijn per definitie positief
- Een score van 1,0 staat voor evenwicht.
- Een score lager dan 1,0 staat voor een relatief lage kans op een stage of leerbaan (de vraag naar stages is bijvoorbeeld hoger dan het aanbod).
- Een score hoger dan 1,0 staat voor een relatief hoge kans op een stage of leerbaan (het aanbod van stages is hoger dan de vraag).

2.3.6 Bereken de derde indicator: de regionale mobiliteit

Voor het afgelopen schooljaar is informatie beschikbaar over (vrijwel) alle BPVOs uit dat schooljaar. Per BPVO is niet alleen informatie beschikbaar over de opleiding (kwalificatie en type leerweg), maar ook over de arbeidsmarktregio waar de leerling woont (staat ingeschreven) en de arbeidsmarktregio waarin het leerbedrijf gevestigd is waar de leerling zijn of haar stage of leerbaan volgt.

Omdat deze informatie voor vrijwel alle BPVOs beschikbaar is, kunnen per opleiding, type leerweg en arbeidsmarktregio zowel de bruto instroom als de bruto uitstroom bepaald worden:

- Bruto uitstroom uit regio x: het aantal studenten (per opleiding en type leerweg) dat in regio x woont maar een stage of leerbaan in een andere regio heeft.
- Bruto instroom in regio x: het aantal studenten (per opleiding en type leerweg) dat niet in regio x woont maar er wel een stage of leerbaan heeft



Hiermee kan vervolgens de netto instroom bepaald worden (het verschil tussen de bruto instroom in regio x en de bruto uitstroom uit regio x).

Door dit (per opleiding en type leerweg) te relateren aan het aantal studenten dat in regio x woont, kan de relatieve netto instroom berekend worden. Deze heeft de volgende kenmerken:

- de relatieve netto instroom ligt per definitie tussen -1 en oneindig¹;
- een waarde tussen -1 en 0 staat voor netto uitstroom;
- een waarde van 0 betekent dat er per saldo geen in- of uitstroom is;
- een positieve waarde staat voor netto uitstroom.

De derde indicator kan dan berekend worden als: relatieve netto instroom + 1. Deze indicator voldoet dan aan dezelfde randvoorwaarden als indicatoren 1 en 2.

Ruimte voor verbetering

De regionale mobiliteit kan berekend worden voor het afgelopen schooljaar. Impliciet wordt aangenomen dat de regionale mobiliteit niet snel zal veranderen, waardoor de regionale mobiliteit van het afgelopen jaar als indicator gebruikt kan worden voor de regionale mobiliteit van het volgende jaar. Bij een volgende editie van de Kans op Stage zou deze aanname onderzocht kunnen worden (door de regionale mobiliteit voor twee of meerdere jaren met elkaar te vergelijken). Indien nodig kan besloten worden om ook hier een prognose voor te modelleren.

2.3.7 Bereken de kans op stage

Na de eerste vijf stappen zijn vier verschillende indicatoren bekend:

- Indicatoren 1 en 2, die beide iets zeggen over de verhouding tussen aanbod en vraag in het lopende schooljaar.
- Indicator 3, die de mate van regionale mobiliteit in het voorafgaande schooljaar weergeeft.
- Indicator 4: de kans op stage die voor 2016 is bepaald.

Kans op stage: primair op landelijk niveau bepaald

De kans op stage wordt bepaald voor bijna 500 verschillende kwalificaties, 2 verschillende leerwegen en 35 arbeidsmarktregio's. Op regionaal niveau kunnen er in theorie dus bijna 35.000 verschillende kansen berekend worden.

Ook al is er veel informatie beschikbaar over vraag en aanbod op dit gedetailleerde niveau (kwalificatie, leerweg en regio), het is niet mogelijk om op dit gedetailleerde niveau betrouwbare uitspraken te doen over (regionale verschillen in) de kans op stage. De kans op stage wordt daarom primair op landelijk niveau bepaald, op basis van de scores voor indicatoren 1, 2 en 4. Voor elk van deze indicatoren wordt (per kwalificatie en leerweg) de score op landelijk niveau bepaald. De kans op stage wordt dan als een gewogen gemiddelde van deze indicatoren berekend.

De keuze voor de gewichten van de verschillende indicatoren is in overleg met SBB bepaald: de kansen op stage zijn op basis van verschillende gewichten bepaald, waarna de herkenbaarheid en plausibiliteit van de verschillende resultaten door twee experts van SBB vergeleken is. Aan het begin van dit proces is besloten om de scores voor kans op stage 2016 niet te gebruiken bij het bepalen van kans op stage 2017. De belangrijkste reden hiervoor is het grote verschil in de manier waarop beide kansen op

¹ De minimumscore wordt behaald als alle studenten die in een bepaalde regio wonen hun stage of leerbaan in een andere regio lopen, en geen enkele leerling uit een andere regio een stage in deze regio volgt. De bruto instroom is dan gelijk aan nul en de bruto uitstroom is gelijk aan het aantal studenten dat in de desbetreffende regio woont. Dan geldt: $\text{netto relatieve instroom} = 0 - \text{bruto_uitstroom} / \text{aantal_woonachtige_studenten} = 0 - 1 = -1$.



stage berekend zijn. De kans voor stage 2017 is uiteindelijk voor 25% door indicator 1 bepaald en voor 75% door indicator 2.

Regionale invulling op basis van regionale mobiliteit

De informatie over de regionale mobiliteit is vervolgens gebruikt om de regionale verschillen in de kans op stage te bepalen. De regionale kans op stage is berekend als een gewogen gemiddelde van de landelijke kans op stage en de regionale mobiliteit, waarbij de regionale mobiliteit voor 20%² meetelt.

Tot slot worden zowel de nationale als de regionale variant van de kans op stage ingedikt in een variabele met vijf antwoordcategorieën:

- gering;
- matig;
- voldoende;
- ruim voldoende;
- goed.

2.4 Kans op stage voor alle opleidingen bepaald

De zojuist beschreven zes stappen zijn voor alle in 2017 geldige kwalificaties toegepast. De verschillende indicatoren en de kans op stage zijn hiermee voor alle kwalificaties berekend.

Voor de kleinste kwalificaties is de betrouwbaarheid en bruikbaarheid van deze indicatoren en de kans op stage niet duidelijk. Voor een opleiding met landelijk 45 studenten zal de regionale mobiliteit bijvoorbeeld van jaar tot jaar sterk kunnen veranderen, afhankelijk van waar de studenten wonen die dat jaar met hun opleiding stoppen en waar de studenten wonen die zich voor het eerst inschrijven. Dergelijke veranderingen vallen niet te voorspellen. Voor deze opleidingen is het bovendien waarschijnlijk dat ze niet of nauwelijks in de leerbedrijvenquête voorkomen. Dit betekent dat er geen correctiefactoren voor berekend kunnen worden.

SBB wil daarom voor de kleinste opleidingen maar een deel van de informatie publiceren. De uiteindelijke dataset met alle indicatoren en de kans op stage (voor alle kwalificaties, leerwegen en regio's) bevat een paar variabelen die SBB kan gebruiken om kleine opleidingen te selecteren:

- Credo_80plus: een indicatorvariabele met de waarde 1 voor alle kwalificaties met minstens 80 ingeschreven studenten (voor BOL en BBL samen);
- Credo_100plus: een indicatorvariabele met de waarde 1 voor alle kwalificaties met minstens 100 ingeschreven studenten (voor BOL en BBL samen);
- Crebo_leerweg_50plus: een indicatorvariabele met de waarde 1 voor alle kwalificaties en leerwegen met minstens 50 ingeschreven studenten (voor BOL of BBL).

² Ook dit gewicht is bepaald door de herkenbaarheid en plausibiliteit van de uitkomsten van verschillende gewichten te laten vergelijken door twee experts van SBB.



3 De gebruikte data

3.1 Registratiegegevens

3.1.1 Aanbod

Het aanbod van stages en leerbanen (BPVOs) bestaat uit twee onderdelen:

- Aanbod dat nog niet is vervuld: vacatures bij erkende leerbedrijven.
- Aanbod dat is vervuld: bestaande (lopende) BPVOs.

Voor elk van deze twee onderdelen is een dataset met registratiegegevens beschikbaar. In tabel 1 worden de gebruikte datasets benoemd, daarna worden ze kort toegelicht.

tabel 1 Kenmerken brondata aanbod

Naam	Inhoud	Waarvoor gebruikt
160916 Bronbestand vacatures stageplaatsen en leerbanen sept 2016.xlsx	<ul style="list-style-type: none">• Aantal in Artus geregistreerde vacatures op 1 september 2016 (het begin van het huidige schooljaar)	<ul style="list-style-type: none">• Aanbod van BPVOs• Afleiden diverse correctiefactoren
20161117, aantal_lopende_bpvos_20152016_per_crebo_met_woonregio_student_v3.xlsx	<ul style="list-style-type: none">• Aantal lopende BPVOs in het afgelopen schooljaar (T)• Gegevens over locatie studenten en leerbedrijven per BPVO• Omvang van de stage (in uren)	<ul style="list-style-type: none">• Aanbod van BPVOs• Regionale mobiliteit• Afleiden diverse correctiefactoren

Bron: Panteia, 2017

Vacatures bij erkende leerbedrijven

Erkende leerbedrijven kunnen vacatures voor stagiaires BOL en BBL op Artus plaatsen. Gegevens uit Artus zijn gebruikt om een overzicht te krijgen van alle openstaande vacatures aan het begin van het huidige schooljaar (schooljaar T+1). Over de omvang van de vacatures (voor hoeveel uur / week en hoeveel weken de vacature bedoeld is) is geen informatie bekend.

Lopende BPVOs

SBB heeft een database met gegevens over alle individuele BPVOs. Uit deze database is een selectie gemaakt met *geaggregeerde* gegevens over alle lopende BPVOs in het schooljaar 2015_2016. Deze selectie resulteert in een databestand waarin per record het totaal aantal lopende BPVOs wordt weergegeven voor elke unieke combinatie van

- crebocode;
- type leerweg;
- woonplaats studenten (arbeidsmarktregio);
- vestigingsregio leerbedrijven (arbeidsmarktregio);
- startdatum van de BPVO;
- einddatum van de BPVO;
- omvang van de BPVO (in uren).

Dit zijn zeer specifieke combinaties. Voor de meeste records in het databestand geldt dan ook dat het aantal lopende BPVOs voor die specifieke combinatie precies 1 is. Er zijn echter een aantal combinaties met meerdere lopende BPVOs.



De meeste waarnemingen in dit bestand hebben betrekking op individuele kwalificaties (de crebocode ligt tussen 25000 en 25999). Sommige waarnemingen betreffen inschrijvingen op dossierniveau (de crebocode ligt tussen 23000 en 23999) of op een nog hoger niveau (de crebocode ligt tussen 79000 en 79999). De inschrijvingen op dossierniveau worden nog gebruikt, de inschrijvingen op hoger niveau (79000 of hoger) niet meer.

3.1.2 Vraag

De vraag naar stages en leerbanen wordt bepaald door het aantal studenten dat ingeschreven staat bij publiek bekostigde Nederlandse MBO opleidingen. Hierover heeft het SBB geaggregeerde data ter beschikking gesteld (zie tabel 2). De meeste waarnemingen in dit bestand hebben betrekking op individuele kwalificaties (de crebocode ligt tussen 25000 en 25999). Sommige waarnemingen betreffen echter inschrijvingen op dossierniveau (de crebocode ligt tussen 23000 en 23999).

Voor sommige correcties is het nodig om per kwalificatiedossier te weten hoe het aantal gediplomeerden in dat dossier verdeeld is over de onderliggende kwalificaties. Dit kan bepaald worden op basis van een bestand met het aantal studenten dat een diploma heeft gehaald (per jaar, kwalificatie(dossier), leerweg en regio) (zie tabel 2).

Om een prognose te kunnen maken van de vraag naar stages en leerbanen in het schooljaar 2016 /2017 wordt gebruikt gemaakt van referentieramingen die door het ministerie van OCW worden gepubliceerd. Het ministerie van OCW heeft hierover een algemeen rapport gepubliceerd alsmede een aantal tabellen met gegevens per onderwijsniveau³. Voor dit onderzoek zijn tabellen met gegevens over het middelbaar beroepsonderwijs gebruikt, waarin het aantal studenten per schooljaar is uitgesplitst naar leerweg en sector (zie tabel 2).

tabel 2 Kenmerken brondata vraag

<i>Naam</i>	<i>Inhoud</i>	<i>Waarvoor gebruikt</i>
20160926, Inschrijvingen_2009_2015_KansOpStage.xlsx	<ul style="list-style-type: none"> Aantal ingeschreven studenten, per schooljaar, crebo, leerweg, niveau en woonplaats leerling. 	<ul style="list-style-type: none"> Huidige vraag naar BPVOs Diverse correctiefactoren
20160926, gediplomeerden_2009_2015_tbv_KansOpStage.xlsx	<ul style="list-style-type: none"> Aantal studenten dat diploma heeft behaald, per schooljaar, crebo, leerweg, niveau en woonplaats leerling. 	<ul style="list-style-type: none"> Huidige vraag naar BPVOs Diverse correctiefactoren
Pubmbo16.ods	<ul style="list-style-type: none"> Referentieramingen OCW, met prognoses van aantal studenten in het MBO, per schooljaar, leerweg en sector* 	<ul style="list-style-type: none"> Prognose vraag naar BPVOs

Bron: Panteia, 2017 * De referentieramingen delen alle kwalificaties in 4 sectoren in (techniek, zorg & welzijn, economie en groen)

3.1.3 Overig

De databestanden die tot nu toe besproken zijn, hebben allemaal betrekking op de vraag naar of het aanbod van stages en leerbanen (BPVOs). Deze informatie moet een aantal bewerkingen ondergaan, waarvoor additionele datasets nodig zijn.

³ Het rapport en de bijbehorende tabellen zijn beschikbaar op de volgende website: <https://www.rijksoverheid.nl/documenten/rapporten/2016/09/20/referentieraming-2016> (geraadpleegd 23 november 2016)



Sommige bestanden bevatten de postcode van leerbedrijven en/of studenten maar niet de bijbehorende indeling in arbeidsmarktregio's. Deze indeling kan alsnog toegevoegd worden op basis van een koppeltabel waarin per postcode de bijbehorende UWV arbeidsmarktregio staat vermeld (zie tabel 3).

De indeling van individuele opleidingen in kwalificaties en kwalificatiedossiers (vastgelegd via crebocodes) wordt jaarlijks aangepast. Bij het bepalen van de kans op stage moet aan het begin van het traject bepaald worden wat op dat moment de geldige kwalificatiestructuur is; dit overzicht geeft aan over welke kwalificaties en kwalificatiedossiers de kans op stage berekend gaat worden (zie tabel 3).

SBB heeft een koppelschema waarin de crebocodes uit voorgaande jaren (voor zover mogelijk) ingedeeld worden volgens de huidige kwalificatiestructuur (zie tabel 3). Dit koppelschema is noodzakelijk om informatie uit voorgaande jaren op een correcte manier aan de huidige kwalificatiestructuur te koppelen.

tabel 3 Kenmerken brondata overig

<i>Naam</i>	<i>Inhoud</i>	<i>Waarvoor gebruikt</i>
20160101 uwv regiotabel.xlsx	<ul style="list-style-type: none"> per postcode, de bijbehorende UWV arbeidsmarktregio 	<ul style="list-style-type: none"> Registratiedata m.b.t. individuele studenten en BPVO's voorzien van bijbehorende arbeidsmarktregio
160914 kwalificatieoverzicht 2016-2017.xlsx	<ul style="list-style-type: none"> Overzicht van de kwalificaties voor het komende schooljaar (naam en crebocode; crebocode tussen 25000 en 25999) Per kwalificatie, het bijbehorende kwalificatiedossiernummer (crebocode tussen 23000 en 23999) 	<ul style="list-style-type: none"> Controle op databestanden met crebocodes (zorg dat alle crebocodes overeenkomen met het huidige kwalificatieoverzicht)
20160926 Crebo HKS en oud naar sectorkamer en marktsegment.xlsx	<ul style="list-style-type: none"> Koppeling tussen crebocodes zoals die in het afgelopen schooljaar zijn gebruikt en crebocodes volgens het huidige kwalificatieoverzicht. 	<ul style="list-style-type: none"> Controle op databestanden met crebocodes (zorg dat alle crebocodes overeenkomen met het huidige kwalificatieoverzicht)

Bron: Panteia, 2017

3.2 Enquête onder erkende leerbedrijven

3.2.1 De onderzoekspopulatie en het steekproefkader

Volgens de gegevens van SBB waren er in juli 2006 bijna 243.000 leerbedrijven erkend. Een deel hiervan is in de praktijk moeilijk te benaderen (emailadres ontbreekt) en wordt daarom niet tot de onderzoekspopulatie gerekend. De onderzoekspopulatie telt daarom iets minder (ruim 220.000) erkende leerbedrijven. Elk van deze leerwerkbedrijven zou in principe volgend kalenderjaar een vacature voor een leerbaan of een stage kunnen hebben.



SBB heeft besloten om niet alle erkende leerbedrijven voor deze enquête te benaderen. De volgende groepen bedrijven zijn vooraf uitgesloten:

- Bedrijven die in de vier maanden voorafgaand aan deze enquête al door SBB gevraagd zijn om aan een ander onderzoek mee te doen. Deze bedrijven zijn uitgesloten om enquêtemoeheid en irritatie zoveel mogelijk te beperken.
- Bedrijven die (volgens Artus) vacatures voor meer dan 15 verschillende kwalificaties hadden: voor deze bedrijven zou het invullen van de vragenlijst te belastend worden.

Verder is het belangrijk om onderscheid te maken tussen bedrijven die op 1 september 2016 één of meerdere vacatures in Artus hadden geregistreerd, en bedrijven die dat niet hadden. Voor bedrijven met geregistreeerde vacatures die aan de enquête meedoen is het namelijk mogelijk om een vergelijking te maken tussen de informatie in Artus over het aantal geregistreeerde vacatures (per kwalificatie en leerweg) aan het begin van het schooljaar 2016 / 2017, en het totaal aantal stages of leerbanen dat het bedrijf in de loop van 2016 / 2017 verwacht aan te kunnen bieden (een van de vragen uit de enquête). Deze vergelijking wordt gebruikt bij het maken van de prognose van het aanbod van stages en leerbanen (zie paragraaf 4.1).

De onderzoekspopulatie voor dit onderzoek is daarom gestratificeerd op basis van de volgende twee 2 kenmerken:

- Het aantal openstaande vacatures in Artus per 1 september 2016:
 - geen vacatures;
 - vacatures voor maximaal 15 crebo's;
 - vacatures voor meer dan 15 crebo's.
- Of bedrijven in de vier maanden voorafgaand aan deze enquête al door SBB gevraagd zijn om aan een ander onderzoek mee te doen.

Het steekproefkader (dat deel van de onderzoekspopulatie waar de steekproef uit getrokken wordt) bestaat uit alle erkende (en geregistreeerde) leerbedrijven die voor niet meer dan 15 kwalificaties erkend zijn en die de vier maanden voorafgaand aan deze enquête niet voor een ander onderzoek van SBB zijn benaderd.

3.2.2 Bepalen van de weegfactoren

De steekproef is aselekt uit het steekproefkader getrokken. Het steekproefkader is echter geen representatieve afspiegeling van de gehele onderzoekspopulatie. Het gevolg hiervan is dat de kans dat een bedrijf in de steekproef zit samen kan hangen met het aantal openstaande vacatures (ondanks het feit dat de steekproef aselekt getrokken is). Dit is het geval als bedrijven die per 1 september een vacature in Artus hadden geregistreerd, in het voorjaar van 2016 vaker door SBB zijn benaderd om aan andere onderzoeken mee te doen. De kans dat deze bedrijven voor de leerbedrijvenenquête zijn uitgenodigd is hierdoor lager. Hiervoor kan (en moet) gecorrigeerd worden, door weegfactoren te berekenen op basis van het per 1 september in Artus geregistreeerde aantal vacatures.

3.2.3 Aantal deelnemende bedrijven

De leerbedrijvenenquête is door bijna 5500 leerbedrijven ingevuld. De eerste twee datasets in tabel 4 bevatten beide de antwoorden van alle deelnemende bedrijven op alle vragen van de leerbedrijvenenquête. Er zijn echter twee belangrijke verschillen tussen beide datasets:

- Beide datasets bevatten een unieke ID-variabele, maar slechts 1 dataset (C11777SBB.sav) bevat daarnaast ook het organisatienummer van de deelnemende bedrijven.
- Elk record in de databestanden bevat de antwoorden van een bedrijf op de verschillende vragen. Omdat elke vraag voor maximaal 15 verschillende opleidingen gesteld kan



worden, bevatten de datasets een aantal variabelen die per bedrijf aangeven op welke specifieke opleidingen de antwoorden betrekking hebben. Beide datasets bevatten daarom de variabelen 'kwal1' t/m 'kwal15'. Dit zijn stringvariabelen die een omschrijving geven van de desbetreffende opleiding. Slechts 1 dataset (KansOpStage_Gekoppeld.sav) bevat daarnaast een aantal variabelen (K1crebo t/m K15crebo) die de crebocode van de desbetreffende opleiding weergeeft. Dit nummer is noodzakelijk om koppelingen met de overige datasets te kunnen maken.

tabel 4 Kenmerken brondata enquête

<i>Naam</i>	<i>Inhoud</i>	<i>Waarvoor gebruikt</i>
KansOpStage_Gekoppeld.sav	<ul style="list-style-type: none"> Resultaten enquête onder erkende leerbedrijven, exclusief organisatierelatienummer 	<ul style="list-style-type: none"> Berekenen diverse correctiefactoren
C11777SBB.sav	<ul style="list-style-type: none"> Resultaten enquête onder erkende leerbedrijven, inclusief organisatierelatienummer 	<ul style="list-style-type: none"> Bepalen organisatierelatienummer van alle leerbedrijven uit de enquête
161026_totale_leerbedrijfbestand_juli.sav	<ul style="list-style-type: none"> Populatiebestand (alle erkende leerbedrijven aan het begin van het huidige schooljaar) 	<ul style="list-style-type: none"> Bepalen weegfactor voor de leerbedrijvenenquête

Bron: Panteia, 2017

Ruimte voor verbetering

Bij het inlezen van de resultaten van de enquête onder leerbedrijven zijn twee verschillende databestanden gebruikt. Beide datasets missen een variabele die noodzakelijk is om koppelingen met overige databestanden te kunnen maken. Hierdoor is het nodig geweest om de resultaten van beide datasets aan elkaar te koppelen.

In de toekomst kunnen betere afspraken met de dataleverancier gemaakt worden, zodat alle benodigde variabelen in één databestand opgeleverd worden. Hierdoor is er voor het verwerken van de resultaten van de leerbedrijvenenquête maar één databestand nodig.

3.2.4 Merendeel deelnemende leerbedrijven niet in Artus

Volgens het bestand met gegevens over vacatures bij erkende leerbedrijven waren er op 1 september 2016 ruim 33.000 erkende leerbedrijven met openstaande vacatures; zij vormen ongeveer 15% van het totaal aantal erkende leerbedrijven. Omdat de steekproef aselekt uit het steekproefkader getrokken is, had een vergelijkbaar laag percentage van de deelnemers aan de leerbedrijvenenquête op 1 september 2016 vacatures openstaan.

Dit heeft zowel een voordeel als een nadeel:

- Een deel van de bedrijven die aan het begin van het schooljaar nog geen vacatures had openstaan, zal in de loop van het schooljaar alsnog vacatures openen. De steekproef bevat veel bedrijven die in het begin van het schooljaar nog geen vacatures hadden geregistreerd. Hierdoor is het mogelijk een betrouwbare inschatting te maken van het aandeel bedrijven uit deze groep dat in de loop van het jaar alsnog vacatures zal openen, en van het aantal vacatures dat deze bedrijven verwacht te openen.
- Een deel van de bedrijven die aan het begin van het schooljaar al vacatures in Artus had geregistreerd, zal in de loop van het schooljaar nieuwe vacatures openen. De steekproef



bevat maar weinig bedrijven die in het begin van het schooljaar al vacatures in Artus hadden geregistreerd. De inschatting van het aandeel bedrijven uit deze groep dat in de loop van het jaar nieuwe vacatures zal openen is hierdoor niet zo betrouwbaar, en dat geldt ook voor het aantal nieuwe vacatures dat deze bedrijven verwachten te openen.

Ruimte voor verbetering

Voor de volgende kans op stage kan overwogen worden om de steekproef anders te stratificeren, door meer bedrijven in de steekproef op te nemen die aan het begin van het schooljaar vacatures in Artus hadden geregistreerd.

3.2.5 Onderwerpen in de enquête

De vragenlijst gaat over de kwalificaties waarvoor het leerbedrijf erkend is. Dit houdt in dat de kwalificaties waarop de vragen gaan, per bedrijf kunnen variëren. Dit is mogelijk omdat per bedrijf bekend is voor welke kwalificaties ze een erkend leerbedrijf zijn. Deze informatie bepaalt ook het aantal vragen dat ze krijgen (een leerbedrijf dat voor 5 kwalificaties erkend is, krijgt ongeveer 5 keer zoveel vragen als een bedrijf dat voor 1 kwalificatie erkend is).

De volgende onderwerpen komen in de vragenlijst aan bod:

1. Per kwalificatie en per leerweg wordt vastgesteld hoeveel mbo-studenten in schooljaar 2015 / 2016 een stage of leerbaan hebben gehad⁴.
2. Per kwalificatie en per leerweg wordt gevraagd, of het bedrijf in schooljaar 2015 / 2016 stages of leerbanen beschikbaar had die niet zijn ingevuld.
3. Per kwalificatie en per leerweg wordt gevraagd, hoeveel stages of leerbanen het bedrijf in schooljaar 2016 / 2017 verwacht aan te kunnen bieden.
4. Voor schooljaar 2016 / 2017 wordt gevraagd hoeveel mbo-studenten het bedrijf tegelijkertijd kan plaatsen.
5. Voor de leerweg BOL wordt per kwalificatie gevraagd, voor hoeveel weken per jaar en voor hoeveel uren per week het bedrijf één stageplaats voor één BOL-student beschikbaar heeft.

⁴ Deze informatie is bekend in registraties van SBB; deze registratiegegevens zijn ter controle aan de bedrijven voorgelegd.



4 Complicaties en correctiefactoren

4.1 Correctiefactor ontbreken leerweg

Voor een deel van de vacatures die in Artus geregistreerd staan is niet bekend om wat voor leerweg het gaat (voor 30% van alle waarnemingen staat 'BOL/BBL' als leerweg vermeld). Deze informatie kan niet geaggregeerd worden naar het niveau van de twee afzonderlijke leerwegen. Op basis van de leerbedrijvenquête wordt daarom een "correctiefactor ontbreken leerweg" berekend. Deze correctiefactor geeft voor elke kwalificatie weer wat het aandeel van het aantal stages (BOL) is in het totale verwacht aanbod van stages en leerbanen (BOL + BBL).

Op basis van deze correctiefactor worden de in Artus geregistreerde vacatures waarvan de leerweg niet bekend is, over de twee leerwegen verdeeld.

4.2 Correctiefactoren inschrijvingen en ontbreken leerbedrijven

Het huidige aanbod van stages en leerbanen wordt bepaald op basis van administratieve gegevens over geregistreerde vacatures en lopende BPVOs (zie paragraaf 2.3.1). Om tot een prognose te komen, moet per kwalificatie, type leerweg en arbeidsmarktregio bepaald worden in welke mate het huidige aanbod zal toenemen. Deze toename kan van twee groepen bedrijven komen:

- bedrijven die aan het begin van het schooljaar al een aanbod van stages en/of leerbanen hadden (in de vorm van lopende BPVOs en/of in Artus geregistreerde vacatures): zij kunnen in de loop van het schooljaar nog meer BPVOs afsluiten en/of vacatures openen;
- bedrijven die aan het begin van het schooljaar nog geen aanbod van stages en/of leerbanen hadden: zij kunnen in de loop van het schooljaar alsnog stages en/of leerbanen aanbieden (in de vorm van getekende BPVOs en/of openstaande vacatures).

Voor elk van deze groepen wordt een aparte correctiefactor berekend die de verhouding tussen het huidige aanbod en het verwachte aanbod weergeeft: voor de eerste groep is dit de correctiefactor inschrijvingen, voor de tweede groep de correctiefactor ontbreken leerbedrijven (zie paragraaf 2.3.2).

De correctiefactor leerbedrijven gaat uit van het huidige aanbod. Dit bestaat uit lopende BPVOs en in Artus geregistreerde openstaande vacatures. Voor de openstaande vacatures verwachten we dat leerbedrijven het aantal beschikbare leerbanen en/of stages soms overschatten en soms onderschatten. Dit kan een aantal oorzaken hebben:

- Voor een deel van de vacatures zal gelden dat de behoefte eraan pas in de loop van het jaar ontstaat of onderkend wordt. Dit resulteert in een onderschatting van het aantal vacatures aan het begin van het schooljaar.
- Een deel van de organisaties zal fouten gemaakt hebben bij het registreren van hun vacatures in Artus. Dit kan zowel in een onder- als overschatting van het aantal vacatures resulteren.
- Een deel van de organisaties heeft een vacature voor een stagiaire, die door stagiaires van meerdere opleidingen vervuld kan worden (bijvoorbeeld timmerman of allround timmerman). Voor beide opleidingen kan dan een vacature in Artus geregistreerd worden. Dit resulteert in een overschatting van het aanbod van stages en leerbanen. Stel, bijvoorbeeld, dat een bedrijf in de loop van het jaar 2 stagiaires wil plaatsen (bijv. 1 stagiair timmerman voor sept / dec 2016 en 1 stagiair timmerman voor maart / juni 2017). Dit bedrijf zou dan in Artus 4 vacatures geregistreerd kunnen hebben (2 opleidingen per stagiair, voor 2 stagiaires).



Om voor deze oorzaken te corrigeren is een "correctiefactor inschrijvingen" nodig. Deze correctiefactor kijkt welke bedrijven uit de leerbedrijvenquête ook in het huidige aanbodbestand zitten (dit zijn bedrijven met lopende BPVOs en/of in Artus geregistreerde vacatures). Voor deze bedrijven wordt de verhouding bepaald tussen het huidige aanbod (op basis van beschikbare administratieve gegevens) en het verwachte aanbod (op basis van de antwoorden op vraag 3 uit de enquête). Deze verhouding is de correctiefactor inschrijvingen.

Een (groot) deel van de erkende leerbedrijven zit niet in het huidige aanbodbestand. Een deel van deze bedrijven kan in de loop van het jaar alsnog stages en/of leerbanen aanbieden. De "correctiefactor ontbreken leerbedrijven" geeft het gemiddelde verwachte aanbod (aantal stages en/of leerbanen) per bedrijf weer, voor de groep bedrijven die niet in het huidige aanbodbestand zitten. Deze correctiefactor kan berekend worden op basis van de informatie van bedrijven uit de leerbedrijvenquête die niet in het huidige aanbodbestand zitten. Door deze correctiefactor te vermenigvuldigen met het totaal aantal bedrijven dat niet in het huidige aanbodbestand zit wordt een inschatting verkregen van het totale aanbod dat nog van deze groep bedrijven verwacht wordt.

4.3 Correctiefactor uren_week

Om de kans op stage te bepalen moet voor het aanbod van stages en leerbanen niet alleen het aantal stages en leerbanen bepaald worden, maar ook de (gemiddelde) duur van die stages en leerbanen. Voor de lopende BPVOs is informatie beschikbaar over de omvang van de BPVO (zowel het totaal aantal weken als het aantal uren per week), maar voor de in Artus geregistreerde vacatures ontbreekt informatie over het aantal uren per week (alleen een start- en einddatum zijn bekend, op basis waarvan het aantal weken bepaald kan worden). Op basis van informatie uit de leerbedrijvenquête kan een inschatting gemaakt worden van het gemiddeld aantal uren per week dat een leerling aan deze stages en leerbanen geacht wordt te werken.

4.4 Correctiefactor omvang BPVOs

Voor indicator 2 wordt een prognose gemaakt van het aanbod van stages en leerbanen, in termen van aantal BPVOs (in plaats van aantal uren). Het uitgangspunt hiervoor is het aantal BPVOs in het voorafgaande schooljaar. Dit wordt vermenigvuldigd met een "groeivoet aantal BPVOs". Deze wordt berekend op basis van de leerbedrijvenquête, door een vergelijking te maken van het aantal stages en leerbanen in het voorafgaande schooljaar (vragen 1 en 2 uit de enquête) en het verwachte aantal stages en leerbanen in het huidige schooljaar (vraag 3 uit de enquête).

4.5 Correctiefactor dossierniveau

Diverse administratieve bestanden bevatten gegevens over stages en leerbanen waarvoor wel het kwalificatiedossier bekend is (crebocode tussen 23000 en 23999) maar niet het juiste kwalificatieniveau (crebocode tussen 25000 en 25999). Bij het bepalen van de huidige en verwachte vraag en aanbod willen we deze informatie wel kunnen gebruiken. Hiervoor is het nodig om, per kwalificatiedossier en leerweg, een verdeelsleutel te maken die weergeeft hoe de vraag of aanbod op dossierniveau verdeeld moet worden over de onderliggende kwalificaties. Hiervoor wordt de "correctiefactor dossierniveau" afgeleid. Hiervoor wordt gebruik gemaakt van informatie over het aantal gediplomeerden: per kwalificatiedossier wordt bepaald hoe



het aantal gediplomeerden van dat dossier verdeeld is over de onderliggende kwalificaties.

4.6 Geen correctie voor toe- en uittreding

Een deel van het aanbod in het huidige schooljaar kan afkomstig zijn van bedrijven die aan het begin van het schooljaar nog niet als leerbedrijf erkend zijn. Met deze bedrijven wordt geen rekening gehouden bij het bepalen van het aanbod van stages en leerbanen (en ook niet bij de verschillende correctiefactoren). Om hier rekening mee te houden zou een "correctiefactor toetreding" afgeleid kunnen worden, die aangeeft met welke fractie het huidige aanbod aangepast moet worden om rekening te houden met nieuwe leerbedrijven.

Een deel van de organisaties verliest in de loop van het schooljaar haar erkenning als leerbedrijf (omdat ze de bedrijfsactiviteiten beëindigen, of om een andere reden). Een deel van deze organisaties kunnen in het huidige aanbodbestand zitten (omdat ze nog lopende BPVOs hebben en/of vacatures in Artus hebben geregistreerd). Dit kan zorgen voor een overschatting van het verwachte aanbod voor het huidige schooljaar. Om hiervoor te corrigeren zou een correctiefactor uittreding afgeleid kunnen worden. Deze correctiefactor zou aan moeten geven met welke fractie het huidige aanbod aangepast moet worden om rekening te houden met het voortijdige vertrek van erkende leerbedrijven.

Beide correctiefactoren zouden in één correctiefactor gecombineerd kunnen worden: een "correctiefactor toe- en uittreding". In overleg met SBB is besloten deze correctiefactor dit jaar nog niet uit te werken.

Ruimte voor verbetering

In de toekomst kan een correctie voor toe- en uittreding van bedrijven toegevoegd worden. Hiervoor is het belangrijk om inzicht te hebben in de maandelijkse mutaties in het aantal erkende leerbedrijven. SBB heeft toegezegd om in het schooljaar 2016 / 2017 maandelijks een momentopname van de gegevens uit Artus te maken, zodat na afloop van dit schooljaar de bedrijvendynamiek van de erkende leerbedrijven (toe- en uittreding) in kaart gebracht kan worden.



5 De methodiek verder uitgewerkt

De berekening van de kans op stage voor 2017 is vastgelegd in 12 verschillende SPSS syntaxen (programma's). Deze syntaxen zijn zo opgezet dat ze in een specifieke volgorde uitgevoerd moeten worden⁵:

- In de eerste syntax (syntax _1_) worden alle te gebruiken datasets ingelezen.
- In de volgende drie syntaxen (syntax_2_ t/m syntax_4_) wordt het huidige aanbod van stages en leerbanen afgeleid. Dit wordt gecombineerd met het afleiden van diverse correctiefactoren.
 - In syntax_2_ wordt een eerste versie van het aanbodbestand gemaakt door de twee databestanden met gegevens over lopende BPVOs en vacatures te combineren. Hierbij worden al een aantal correcties uitgevoerd, waarvoor informatie uit de enquête gebruikt wordt (voor deze correcties zijn nog geen weegfactoren nodig).
 - Voor de meeste correcties geldt dat gewogen resultaten van de enquête nodig zijn. De weegfactoren voor de enquête worden in syntax_3_ bepaald. In syntax_4_ worden vervolgens alle analyses op de enquête uitgevoerd die nodig zijn voor de verschillende correcties. In syntax_4_ wordt ook het huidige aanbodbestand van stages en leerbanen afgeleid.
- In syntax_5_ worden voor zowel de afgesloten BPVOs als de inschrijvingen en gediplomeerden de aantallen per kwalificatie, leerweg en/of regio bepaald. De resultaten worden uitgesplitst naar kwalificatieniveau: er worden aparte datasets opgeslagen met informatie op kwalificatieniveau en op kwalificatiedossierniveau.
- Syntax_6_ begint met het bepalen van de huidige vraag naar stages en leerbanen: per kwalificatie, leerweg en regio wordt het aantal inschrijvingen in het afgelopen schooljaar bepaald, door inschrijvingen op kwalificatieniveau te combineren met inschrijven op kwalificatiedossierniveau.
- In syntax_7_ en syntax_8_ wordt de prognose van het aanbod van stages en leerbanen (ten behoeve van indicator 1) gemaakt. Eerst wordt in syntax_7_ de omvang van stages en leerbanen bepaald (uren per week en weken per jaar), vervolgens worden in syntax_8_ verschillende bestanden gecombineerd om de prognose van het aanbod van stages en leerbanen te bepalen.
- In syntax_9_ en syntax_10_ worden de huidige en verwachte vraag naar stages en leerbanen afgeleid.
- In syntax_11_ wordt de regionale mobiliteit bepaald.
- De eerder afgeleide bestanden worden in syntax_12_ gebruikt om de vier indicatoren te bepalen. Vervolgens wordt de kans op stage berekend, op basis van de scores op deze 4 indicatoren. Aan het eind van deze syntax wordt een definitief databestand opgeslagen met daarin per crebo, leerweg en arbeidsmarktregio de scores op de vier indicatoren, de kans op stage, en diverse kenmerken die gebruikt kunnen worden om op de website van Kans op Stage extra achtergrondinformatie te geven.

5.1 _1_hernoem.sps

In deze syntax worden alle te gebruiken datasets ingelezen en krijgen ze een nieuwe, interne naam. Deze interne naam wordt gebruikt in alle volgende syntaxen.

Als de kans op stage in de toekomst opnieuw berekend moet worden, op basis van nieuwe datasets, dan hoeven de namen van de nieuwe datasets alleen maar in deze syntax aangepast te worden. De interne namen van de gebruikte datasets blijven immers hetzelfde.

⁵ Naast deze syntaxen zijn er nog een aantal syntaxen waarin de labels van de verschillende opleidingen en arbeidsmarktregio's worden vastgelegd en de koppelingen tussen CREBO-indelingen van verschillende jaren en verschillende niveaus (kwalificatie, kwalificatiedossier, sector, sectorkamer, sectorunit en marktsegment). Deze worden hier niet besproken.



Bij het inlezen van de verschillende datasets worden de volgende correcties en controles uitgevoerd:

- veel databronnen zijn in de vorm van Excel bestanden; deze worden omgezet in SPSS datasets;
- alle variabelen worden van de juiste labels voorzien (deze labels kunnen jaarlijks wijzigen, bijvoorbeeld bij verschillende kwalificaties en/of kwalificatiedossiers);
- bij variabelen die in meerdere datasets voorkomen wordt ervoor gezorgd dat ze overal dezelfde naam krijgen;
- de meetniveaus van variabelen worden gecontroleerd; vaak is het nodig om variabelen die in de brondata als stringvariabele (tekst) zijn gedefinieerd om te zetten naar numerieke variabelen;
- als in de toekomst blijkt dat de namen van sommige variabelen in de brondata zijn veranderd, dan zal dat in deze stap worden aangepast (alle relevante variabelen moeten elk jaar dezelfde naam hebben).

5.1.1 Ruimte voor verbetering

Bij het inlezen van de resultaten van de enquête onder leerbedrijven zijn twee verschillende databestanden gebruikt. Beide datasets missen een variabele die noodzakelijk is om koppelingen met overige databestanden te kunnen maken. Hierdoor is het nodig geweest om de resultaten van beide datasets aan elkaar te koppelen.

In de toekomst kunnen betere afspraken met de dataleverancier gemaakt worden, zodat alle benodigde variabelen in één databestand opgeleverd worden. Hierdoor is er voor het verwerken van de resultaten van de leerbedrijvenenquête maar één databestand nodig (deze verbeteringsuggestie is ook in paragraaf 3.2.3 vermeld).

Een andere verbetering is dat bij het inlezen van de verschillende datasets informatie over de bijbehorende sectorkamer toegevoegd kan worden. Dit gebeurt nu vaak in de volgende syntaxis, en vaak wordt het dubbel gedaan (dat kan voorkomen worden door het in deze syntax op te nemen).

5.2 **_2_brongegevens vacatures corrigeren.sps**

In deze syntax wordt de informatie over openstaande vacatures en lopende BPVOs gecombineerd tot één bestand. Dit vormt een eerste versie van het aanbodbestand.

Voordat deze bestanden gekoppeld kunnen worden, moeten een aantal correcties worden doorgevoerd. Voor sommige van deze correcties moeten gegevens van de leerbedrijvenenquête gebruikt worden. Deze syntax begint dan ook met het inlezen van de leerbedrijvenenquête.

5.2.1 Leerbedrijvenenquête bewerken

Om de gegevens van de verschillende bestanden te kunnen combineren, moeten ze allemaal op dezelfde manier gestructureerd zijn: een record moet informatie bevatten over de lopende BPVOs of openstaande vacatures per leerbedrijf, kwalificatie (crebocode) en leerweg (BOL of BBL). Als leerbedrijven BPVOs of vacatures voor meerdere kwalificaties en/of leerwegen hebben, dan zijn er meerdere records voor deze leerbedrijven in de dataset.

De datasets met informatie over openstaande vacatures en lopende BPVOs zijn al op deze manier gestructureerd. De dataset met de resultaten van de leerbedrijvenenquête is echter anders gestructureerd: deze dataset 1 record per leerbedrijf, ongeacht het aantal kwalificaties en/of leerwegen waarover vragen aan dit bedrijf gesteld zijn. De



structuur van de dataset met de resultaten van de leerbedrijvenquête wordt daarom aangepast, zodat het overeenkomt met de structuur voor de overige datasets.

5.2.2 *Dataset openstaande vacatures inlezen*

Vervolgens wordt het databestand met informatie over openstaande vacatures ingelezen. Hierop worden enkele controles uitgevoerd.

Een van deze controles is of er per leerbedrijf, kwalificatie (crebocode) en type leerweg maar 1 record in de dataset voorkomt. Dit blijkt meestal het geval te zijn, maar niet altijd: iets minder dan 2% van de waarnemingen betreft een dubbeltelling. In de helft van de gevallen is er sprake van verschillen in startdatum, einddatum en/of het aantal vacatures, wat suggereert dat de verschillende records verschillende vacatures betreffen. In deze gevallen wordt het aantal vacatures in de betrokken records opgeteld, om zo alsnog per leerbedrijf, kwalificatie (crebocode) en type leerweg 1 record in de dataset te hebben dat het totaal aantal openstaande vacatures weergeeft. In de andere gevallen (waarbij er 2 of meer records zijn met dezelfde startdatum, einddatum en aantal openstaande vacatures) hebben we aangenomen dat er sprake is van een dubbeltelling en zijn de dubbele records verwijderd (zodat er opnieuw maar 1 record overblijft).

5.2.3 *Arbeidsmarktregio bepalen*

De volgende stap is dat per (vestiging van een) leerbedrijf wordt bepaald in welke arbeidsmarktregio het gevestigd is. Dit gebeurt op basis van de postcode.

5.2.4 *Correctie onbekende leerweg*

Voor bijna een kwart van de in Artus geregistreerde vacatures is het type leerweg niet bekend. Dit kan deels opgelost worden door het bestand met geregistreerde vacatures te koppelen aan de resultaten van de leerbedrijvenquête. Immers, beide bestanden vormen een deel van de populatie van erkende leerbedrijven, en beide bestanden bevatten het organisatierelatienummer op basis waarvan ze geïdentificeerd kunnen worden. Omdat in de leerbedrijvenquête vragen over openstaande vacatures gesteld zijn (inclusief de desbetreffende leerweg) is het mogelijk om de in Artus ontbrekende informatie over het type leerweg alsnog toe te voegen.

Dit is echter alleen mogelijk voor die leerbedrijven die in beide bestanden voorkomen. De overlap tussen beide bestanden is echter klein (zie paragraaf 3.2.4). Hierdoor kan deze correctie maar voor weinig leerbedrijven gemaakt worden. Voor ongeveer 300 records uit Artus waarvoor de leerweg niet bekend was konden we op deze manier bepalen of het om BOL of BBL ging; dit is een fractie van het totaal aantal records waarvoor de leerweg in Artus niet bekend is.

5.2.5 *Voeg gegevens over lopende BPVOs toe*

Het bestand met (gecorrigeerde) gegevens over openstaande vacatures wordt uitgebreid met het bestand met gegevens over lopende BPVOs. Dit zijn alle waarnemingen uit het bestand met alle BPVOs uit het vorige schooljaar die aan de volgende eisen voldoen:

- ze zijn in Nederland gerealiseerd;
- ze hebben betrekking op een BOL of BBL traject;
- de kwalificatie en/of het kwalificatiedossier is bekend (de crebocode is aanwezig en ligt tussen 23000 en 25999);



- twee weken na het begin van het huidige schooljaar⁶ lopen ze nog.

Ongeveer 65.000 records uit het bestand met lopende BPVOs van het vorige schooljaar voldoen aan deze eisen (80% BBL, 20% BOL).

Dit onderdeel van de syntax bevat verwijzingen naar specifieke jaartallen. Deze verwijzingen moeten elk jaar worden aangepast.

5.2.6 *Creëer eerste versie aanbodbestand*

Aan het einde van `syntax_2_` is een dataset gecreëerd die per leerbedrijf, kwalificatie(dossier) en type leerweg maximaal twee records kan bevatten: een record met het aantal nog niet gerealiseerde BPVOs (aantal openstaande vacatures) en een record met het aantal gerealiseerde BPVOs (aantal lopende BPVOs aan het begin van het schooljaar). Voor het bepalen van het totale aanbod maken we geen onderscheid tussen wel of niet gerealiseerde BPVOs. Aan het einde van deze syntax worden de openstaande vacatures en lopende BPVOs daarom (per leerbedrijf, kwalificatie(dossier) en type leerweg) bij elkaar opgeteld.

Hiermee ontstaat een bestand dat per leerbedrijf, kwalificatie(dossier) en type leerweg één record heeft met daarin het aanbod (in termen van BPVOs). Dit is de eerste versie van het aanbodbestand.

5.3 **`_3_` weegfactoren leerbedrijvenquête.sps**

Het doel van deze syntax is om de juiste weegfactoren af te leiden voor de leerbedrijvenquête. Dit gebeurt door een vergelijking te maken tussen het aantal bedrijven dat de enquête heeft ingevuld en het aantal bedrijven in de onderzoekspopulatie. Hierbij moet rekening gehouden worden met de stratificatie van de steekproef.

5.3.1 *De stratificatie van de populatie*

De onderzoekspopulatie voor dit onderzoek is gestratificeerd op basis van de volgende twee 2 kenmerken:

- Het aantal openstaande vacatures in Artus per 1 september 2016:
 - geen vacatures;
 - vacatures voor maximaal 15 crebo's;
 - vacatures voor meer dan 15 crebo's.
- Of bedrijven in de vier maanden voorafgaand aan deze enquête al door SBB gevraagd zijn om aan een ander onderzoek mee te doen.

Het startpunt is een bestand met een overzicht van alle door SBB geregistreerde en erkende leerbedrijven (peildatum: juli 2016). Dit bestand bevat bijna 243.000 leerbedrijven. Een deel hiervan is in de praktijk moeilijk te benaderen (emailadres ontbreekt) en wordt daarom niet tot de onderzoekspopulatie gerekend. Na verwijdering van deze bedrijven blijft een populatiebestand over met ruim 220.000 erkende leerbedrijven.

Om dit populatiebestand te kunnen stratificeren, moet per bedrijf bepaald worden of ze per 1 september 2016 vacatures in Artus hadden geregistreerd (en zo ja, hoeveel). Hiervoor is het populatiebestand gekoppeld aan het bronbestand met vacatures voor stageplaatsen en leerbanen (zie tabel 1).

⁶ Strikt genomen hadden we de grens bij het begin van het huidige schooljaar kunnen leggen en niet twee weken daarna. Het risico hiervan is dat trajecten die vrijwel geheel in het vorige schooljaar speelden maar door omstandigheden pas in de eerste (of tweede) week van het nieuwe schooljaar zijn afgerond worden meegeteld.



Een deel van de bedrijven uit het bronbestand (en zelfs een paar bedrijven die aan de enquête hebben meegedaan) komen niet voor in het bestand met geregistreerde en erkende leerbedrijven. Misschien komt dit omdat er een paar maanden zit tussen de peildatum waarop het bestand met geregistreerde en erkende leerbedrijven is bepaald, de peildatum waarop het aantal geregistreerde vacatures in Artus is bepaald en de periode waarin de enquête is gehouden.

Naast het aantal openstaande vacatures moet ook bepaald worden of bedrijven eerder door SBB gevraagd zijn om aan een ander onderzoek mee te doen. Informatie hierover is aanwezig in het populatiebestand met alle door SBB geregistreerde en erkende leerbedrijven: de variabele 'uitsluiten_onderzoek' geeft weer of een bedrijf om deze reden van de enquête uitgesloten moest worden of niet.

Op basis van deze informatie kan het populatiebestand gestratificeerd worden, waarna het aantal bedrijven per stratum bepaald kan worden.

5.3.2 Bepalen van de weegfactoren

Om de weegfactoren voor de steekproef te bepalen wordt de steekproef op dezelfde manier gestratificeerd als het populatiebestand, waarna per stratum een vergelijking gemaakt kan worden tussen de relatieve omvang van dat stratum in de steekproef en in de populatie.

5.4 _4_leerbedrijvenenquête.sps

In deze syntax worden de (gewogen) resultaten van de leerbedrijvenenquête gebruikt om diverse correctiefactoren te berekenen. Deze worden in diverse (tijdelijke) datasets opgeslagen die later toegevoegd kunnen worden aan databestanden met informatie over (vooral) het aanbod van stages en leerbanen.

5.4.1 Correctiefactor ontbreken leerweg

In vraag drie van de enquête wordt aan bedrijven gevraagd of ze, per kwalificatie en per leerweg, kunnen aangeven hoeveel stages of leerbanen ze voor het huidige schooljaar (2016 / 2017) verwachten aan te kunnen bieden.

Door de antwoorden op deze vraag te aggregeren naar het niveau van kwalificaties en leerwegen kan per kwalificatie en per leerweg het verwachte aanbod voor het huidige schooljaar bepaald worden. De correctiefactor ontbreken leerweg wordt dan berekend als het aandeel verwachte stageplaatsen (BOL) in het totale verwachte aanbod (BOL + BBL).

Voor kwalificaties met weinig studenten is het niet mogelijk om de correctiefactor op het niveau van de kwalificaties te berekenen (de enquête bevat te weinig waarnemingen voor deze kwalificatie). Voor deze kwalificaties wordt de correctiefactor op het niveau van kwalificatiedossiers bepaald.

Vervolgens worden deze correctiefactoren gebruikt om in de eerste versie van het aanbodbestand de aantallen openstaande vacatures (per kwalificatie en regio) waarvoor de leerweg niet bekend is, over BOL en BBL te verdelen. Hierna wordt een gecorrigeerde (en definitieve) versie van het huidige aanbodbestand opgeslagen. Voor een andere correctie is een versie van dit bestand aangemaakt en opgeslagen die het huidige aanbod weergeeft voor leerbedrijven die voor maximaal 15 verschillende kwalificaties vacatures in Artus geregistreerd hebben.



Ruimte voor verbetering

In de definitieve versie is geen gebruik gemaakt van het databestand met het huidige aanbod voor leerbedrijven die voor maximaal 15 verschillende kwalificaties vacatures in Artus geregistreerd hebben. Bij een volgende versie kan dit deel van de syntax verwijderd worden.

5.4.2 Correctiefactor inschrijvingen

Deze correctiefactor wordt bepaald op basis van bedrijven die in het huidige aanbodbestand zitten en aan de leerbedrijvenquête hebben meegedaan. Voor zover deze bedrijven aan het begin van het huidige schooljaar vacatures in Artus hadden geregistreerd, kunnen we enkel gegevens gebruiken van bedrijven die hierbij aangeven op welk type leerweg deze vacatures betrekking hebben (deze informatie is noodzakelijk om de correctiefactor per leerweg te kunnen bepalen).

Als uitgangspunt voor deze correctiefactor wordt daarom de eerste versie van het aanbodbestand gebruikt (in deze versie heeft de correctie voor het ontbreken van de leerweg nog niet plaatsgevonden). Dit bestand wordt gekoppeld aan de leerbedrijvenquête. Vervolgens wordt een selectie gemaakt van bedrijven die in beide bestanden voorkomen.

Voor deze selectie van bedrijven wordt (per leerweg) het huidige aanbod (lopende BPVO's en openstaande vacatures voor alle kwalificaties) vergeleken met het verwachte aanbod. Voor het verwachte aanbod worden de antwoorden op vraag 3 uit de leerbedrijvenquête gebruikt. Deze vraag is helaas niet goed ingevuld; slechts voor een kwart van de waarnemingen geldt dat de bedrijven vraag 3 voor alle kwalificaties hebben ingevuld. In combinatie met de relatief hoge uitval van waarnemingen uit Artus (voor plm. 30% van de in Artus geregistreerde vacatures is de leerweg niet bekend) zorgt dit ervoor dat het aantal waarnemingen waar deze correctiefactor op gebaseerd kan worden erg beperkt is. De correctiefactor kan hierdoor alleen naar type leerweg (BOL of BBL) worden uitgesplitst (zelfs voor een uitsplitsing naar leerweg en sectorkamer is het aantal waarnemingen te laag).

De correctiefactor inschrijvingen wordt berekend door (per leerweg) het totale verwachte aanbod (aantal verwachte stages of leerbanen) te delen door het totale huidige aanbod (aantal openstaande vacatures en lopende BPVO's).

Ruimte voor verbetering

Als er meer waarnemingen beschikbaar zijn, dan zou deze correctiefactor op een meer gedetailleerd niveau berekend kunnen worden (bijvoorbeeld naar leerweg en sectorkamer). Het aantal waarnemingen kan op een aantal manieren verhoogd worden:

- Verbeter de registratie in Artus: verlaag het aandeel bedrijven dat bij een vacature niet aangeeft om welk type leerweg het gaat.
- Pas de stratificatie van de steekproef voor de leerbedrijvenquête zo aan, dat er meer bedrijven uitgevraagd worden die vacatures in Artus hebben openstaan (dit is ook in paragraaf 3.2.4 vermeld)
- Probeer de respons op vraag 3 in de enquête te verhogen.

5.4.3 Correctiefactor ontbreken leerbedrijven

Ook voor deze correctiefactor wordt de eerste versie van het aanbodbestand (zonder correctie voor het ontbreken van de leerweg) gekoppeld aan de leerbedrijvenquête.



Vervolgens wordt een selectie gemaakt van bedrijven die in de leerbedrijvenquête zitten maar niet in het aanbodbestand.

Voor deze groep bedrijven wordt, per leerweg en sectorkamer, het gemiddeld aantal stages of leerbanen bepaald dat ze in de loop van het huidige schooljaar verwachten (op basis van vraag 3).

Belangrijk hierbij is dat het gemiddelde berekend wordt ten opzichte van alle erkende leerbedrijven in de leerbedrijvenquête (ongeacht sectorkamer en leerweg), en niet alleen maar de leerbedrijven die voor (opleidingen uit) de desbetreffende sectorkamer als leerbedrijf erkend zijn. Dit heeft te maken met hoe deze correctiefactor gebruikt moet worden. Het doel van deze correctiefactor is om een prognose te maken van het aanbod van stages en leerbanen bij erkende leerbedrijven die niet in het huidige aanbodbestand zitten. Om deze prognose te maken zal de te berekenen correctiefactor vermenigvuldigd worden met het aantal erkende leerbedrijven dat niet in het huidige aanbodbestand zit. In dit aantal zitten alle erkende leerbedrijven, zonder onderscheid naar kwalificatie, kwalificatiedossier of type leerweg. Dit betekent dat de te berekenen correctiefactor berekend moet worden als een gemiddelde ten opzichte van alle leerbedrijven in de leerbedrijvenquête, zonder onderscheid naar kwalificatie, kwalificatiedossier of type leerweg.

De correctiefactor ontbreken leerbedrijven wordt vervolgens vermenigvuldigd met het totaal aantal erkende leerbedrijven dat niet in het aanbodbestand zit. Het resultaat (een prognose van het aanbod van stages en leerbanen afkomstig van leerbedrijven die niet in het huidige aanbodbestand zitten) wordt als een apart databestand opgeslagen.

5.4.4 Correctiefactor uren_week

Voor openstaande vacatures voor BOL stages moet een inschatting gemaakt worden van het gemiddeld aantal uur per week dat stagiaires geacht worden te werken (voor BBL leerplaatsen is dit niet nodig). Dit gebeurt op basis van de antwoorden op vraag 5 uit de leerbedrijvenquête.

Sommige antwoorden op deze vraag zijn erg hoog of erg laag; antwoorden die suggereren dat studenten meer dan 40 uur per week of minder dan 2 uur per week aan een BOL stage moeten besteden, worden niet gebruikt bij het berekenen van deze correctiefactor.

Vervolgens wordt het gemiddeld aantal uren per week bepaald dat een BOL stage zou duren. Dit gebeurt op verschillende aggregatieniveaus: per kwalificatie, per kwalificatiedossier, per marktsegment en per sectorkamer.

De uiteindelijke correctiefactor per kwalificatie wordt bepaald door het aantal waarnemingen:

- als er voor een kwalificatie minstens 100 waarnemingen zijn, dan wordt de correctiefactor op kwalificatieniveau gebruikt;
- als dit niet het geval is, maar voor het bijbehorende kwalificatiedossier minstens 100 waarnemingen zijn, dan wordt de correctiefactor op kwalificatiedossierniveau gebruikt;
- als dit niet het geval is, maar voor het bijbehorende marktsegment minstens 100 waarnemingen zijn, dan wordt de correctiefactor op marktsegmentniveau gebruikt;
- als dit niet het geval is, maar voor de bijbehorende sectorkamer minstens 100 waarnemingen zijn, dan wordt de correctiefactor op sectorkamerniveau gebruikt;



- als er ook op het niveau van de sectorkamer minder dan 100 waarnemingen zijn, dan wordt het gemiddelde voor alle BOL stages gebruikt.

De resulterende correctiefactor wordt in een apart databestand opgeslagen. In syntax_8_ wordt deze correctiefactor vervolgens gebruikt.

5.4.5 Groeivoet aantal BPVOs

Het aanbod van stages en leerbanen is de optelsom van vervuld aanbod (afgesloten BPVOs) en onvervuld aanbod (openstaande vacatures). Voor het voorafgaande schooljaar zijn hierover twee vragen in de enquête gesteld (vragen 1 en 2). Over het verwachte aantal stages en leerbanen voor het huidige schooljaar is 1 vraag in de enquête opgenomen (vraag 3). Voor dit verwachte aantal plaatsen is niet duidelijk of deze allemaal vervuld zullen worden of niet. Daarom is de groeivoet bepaald op basis de verhouding tussen de antwoorden op vraag 3 enerzijds, en de antwoorden op vragen 1 en 2 anderzijds.

Het al dan niet opnemen van vraag 2 maakt veel uit voor de resulterende groeivoet: als vraag 3 enkel met vraag 1 vergeleken wordt dan is de resulterende groeivoet veel hoger dan wanneer de antwoorden van vragen 1 en 2 bij elkaar opgeteld worden.

De groeivoet wordt per leerweg berekend als: het verschil tussen het verwachte aanbod (vraag 3) en het aanbod in het voorafgaande jaar (vragen 1 en 2), gedeeld door het aanbod in het voorafgaande jaar (vragen 1 en 2)⁷. Dit gebeurt op verschillende aggregatieniveaus: per kwalificatie, per kwalificatiedossier, per marktsegment en voor alle opleidingen.

De uiteindelijke groeivoet per kwalificatie wordt bepaald door het aantal waarnemingen:

- als er voor een kwalificatie minstens 55 waarnemingen zijn, dan wordt de correctiefactor op kwalificatieniveau gebruikt;
- als dit niet het geval is, maar voor het bijbehorende kwalificatiedossier minstens 55 waarnemingen zijn, dan wordt de correctiefactor op kwalificatiedossierniveau gebruikt;
- als dit niet het geval is, maar voor het bijbehorende marktsegment minstens 55 waarnemingen zijn, dan wordt de correctiefactor op marktsegmentniveau gebruikt;
- als er ook op het niveau van het marktsegment minder dan 55 waarnemingen zijn, dan wordt het gemiddelde voor alle opleidingen (per type leerweg) gebruikt.

Bijzonder hoge en lage groeivoeten zijn vervolgens aangepast: groeivoeten van meer dan 50% zijn op 50% gesteld, groeivoeten van minder dan -50% (halvering van het aanbod) zijn op -50% gesteld.

Om de te hanteren grenswaarde te kunnen bepalen, zijn indicatoren 1 en 2 (en de uiteindelijke kans op stage) berekend op basis van verschillende grenswaarden. Vervolgens zijn de resulterende waarden voor deze indicatoren (en de kans op stage) voor een selectie van opleidingen met elkaar vergeleken. Op basis van deze vergelijking is in overleg met SBB besloten om uit te gaan van een grenswaarde van 55 waarnemingen. Ook de afkapperpercentages voor bijzonder hoge en lage groeivoeten zijn in overleg met SBB gekozen.

De resulterende correctiefactor wordt in een apart databestand opgeslagen.

⁷ In formulevorm: $(\text{verwacht aanbod} - \text{aanbod voorafgaand jaar}) / (\text{aanbod voorafgaand jaar})$. Dit is equivalent met $(\text{verwacht aanbod}) / (\text{aanbod voorafgaand jaar}) - 1$.



5.5 *_5_correctiefactor dossier.sps*

5.5.1 *Aantal afgesloten BPVOs*

Het bestand met aantal lopende BPVOs in het afgelopen schooljaar wordt geopend. Vervolgens wordt per kwalificatie, type leerweg en regio het totaal aantal afgesloten BPVOs bepaald. Deze informatie wordt in `syntax_9_` gebruikt bij het bepalen van het gemiddeld aantal BPVOs per ingeschreven leerling, en in `syntax_12_` voor het bepalen van de huidige vraag naar BPVOs.

Ruimte voor verbetering

De informatie over aantallen BPVOs in het afgelopen schooljaar is gebaseerd op gegevens over afgesloten BPVOs waarvoor het kwalificatieniveau en de woonregio van de student bekend zijn. Voor ongeveer 5% van alle BPVOs ontbreekt deze informatie. Deze BPVOs zijn niet meegeteld bij het bepalen van het aantal BPVOs in het afgelopen schooljaar. Dit kan voor een kleine onderschatting van het totaal aantal BPVOs zorgen. Dit kan eventueel ondervangen worden door een correctiefactor te berekenen die de mate van onderschatting weergeeft. Het liefst op een zo gedesaggregeerd niveau mogelijk, maar als de crebocode ontbreekt dan kan dit lastig blijken.

5.5.2 *Aantal ingeschreven studenten*

Het bestand met jaarlijkse informatie over ingeschreven studenten wordt geopend. Hieruit wordt een selectie gemaakt van alle studenten die in het afgelopen schooljaar (2015 / 2016) ingeschreven waren.

Per kwalificatiedossier (crebocode tussen 23000 en 23999) wordt nu bepaald hoeveel studenten er ingeschreven stonden (in totaal, per leerweg en per leerweg x regio). Hetzelfde gebeurt op het niveau van kwalificaties (crebocode tussen 25000 en 25999).

Deze informatie wordt niet gebruikt om de correctiefactor dossierinschrijvingen te bepalen.

5.5.3 *Aantal gediplomeerden*

Het bestand met jaarlijkse informatie over aantal gediplomeerden op kwalificatieniveau wordt geopend. Hieruit wordt een selectie gemaakt van alle studenten die in het meest recent beschikbare schooljaar (2014 / 2015) een diploma gehaald hebben.

Per kwalificatiedossier (crebocode tussen 23000 en 23999) wordt nu bepaald hoeveel studenten een diploma gehaald hebben (in totaal, per leerweg en per leerweg x regio). Hetzelfde gebeurt op het niveau van kwalificaties (crebocode tussen 25000 en 25999).

5.5.4 *Splitsen datasets*

In de voorgaande drie stappen zijn drie verschillende datasets gemaakt, met informatie over aantallen afgesloten BPVOs, aantallen studenten en aantallen gediplomeerden.

Elk van deze datasets wordt nu in twee delen opgesplitst:

- een deel met alle waarnemingen op kwalificatieniveau (crebocode tussen 25000 en 25999);
- een deel met alle waarnemingen op kwalificatiedossierniveau (crebocode tussen 23000 en 23999).



5.5.5 Berekenen correctiefactor dossierniveau

De dataset met informatie over aantallen gediplomeerden op kwalificatieniveau wordt ingelezen. De correctiefactor dossierniveau wordt vervolgens per kwalificatie en leerweg berekend als de verhouding tussen:

- het aantal studenten dat in dit schooljaar een diploma heeft gehaald voor deze *kwalificatie* en leerweg;
- het aantal studenten dat in dit schooljaar een diploma heeft gehaald voor het bovenliggende *kwalificatiedossier* en leerweg.

De resulterende correctiefactor wordt in een apart databestand opgeslagen.

5.6 **_6_aantal_inschrijvingen.sps**

De berekende correctiefactor dossierniveau wordt in deze syntax gebruikt om inschrijvingen op kwalificatiedossierniveau over de onderliggende kwalificaties te verdelen.

Uitgangspunt hiervoor is het bestand met alle inschrijvingen op kwalificatieniveau. Door alle inschrijvingen te aggregeren naar kwalificatiedossierniveau wordt bepaald hoeveel inschrijvingen er per kwalificatiedossier zijn waarvoor het kwalificatieniveau al bekend is. Dit totaal is dus exclusief de inschrijvingen op dossierniveau. Dit bestand bevat echter ook een variabele die het totaal aantal inschrijvingen per kwalificatiedossier bevat, *inclusief de inschrijvingen op dossierniveau* (dit is in `syntax_5_` gedaan). Het verschil tussen deze twee totalen (het totaal inclusief en exclusief inschrijvingen op dossierniveau) is precies het aantal inschrijvingen op dossierniveau, dat over de onderliggende kwalificaties verdeeld moet worden.

Hiervoor kan de correctiefactor dossierniveau gebruikt worden. Deze correctiefactor geeft per kwalificatie en leerweg weer welk deel van de dossierinschrijvingen aan deze kwalificatie en leerweg moet worden toegerekend. De dataset met inschrijvingen is echter op een lager aggregatieniveau gedefinieerd: resultaten zijn niet alleen naar kwalificatie en leerweg, maar ook nog naar regio verbijzonderd. Bij het toerekenen van de dossierinschrijvingen aan de inschrijvingen (per kwalificatie, leerweg en regio) moet dus ook een verdeelsleutel gebruikt worden om de aan een kwalificatie en leerweg toegekende inschrijvingen over de regio's te verdelen. Hiervoor gebruiken we, per kwalificatie en leerweg, de verdeling van het aandeel inschrijvingen over de regio's.

Voor de verdeling van de dossierinschrijvingen over de onderliggende kwalificaties worden zodoende (per kwalificatie, leerweg en regio) de volgende drie variabelen met elkaar vermenigvuldigd:

- het aantal inschrijvingen op dossierniveau
- de correctiefactor dossierniveau
- per kwalificatie en leerweg, het aandeel inschrijvingen per regio.

Aan het einde van deze syntax kan bepaald worden voor welke kwalificaties er minstens 80 inschrijvingen zijn en voor welke minstens 100 (dat kon niet eerder, omdat nu pas de inschrijvingen op dossierniveau aan de onderliggende kwalificaties zijn toegekend).



5.7 **_7_omvang BPVOs.sps**

In deze syntax wordt de omvang van BPVOs in het afgelopen schooljaar bepaald: het gemiddeld aantal uren per week en het gemiddeld aantal weken per jaar van stages en leerbanen waarvoor in het afgelopen schooljaar BPVOs zijn afgesloten.

Uitgangspunt hiervoor is het (in syntaxen `_1_` en `_2_` aangepaste) bestand met informatie over alle afgesloten BPVOs. Dit bestand bevat informatie over onder andere de startdatum, de geplande en werkelijke einddatum, en de omvang van de stage in uren.

De geplande en werkelijke duur van de stageperiode wordt bepaald door het aantal dagen te bepalen tussen de startdatum van de BPVO enerzijds, en de geplande en werkelijke einddatum anderzijds. Dit is inclusief weekenden en mogelijke zon- en feestdagen. We corrigeren niet voor mogelijke zon- en feestdagen, maar wel voor weekenden. Dit doen we door het aantal geplande dagen te vermenigvuldigen met 5/7. Dit levert een inschatting op van het aantal werkdagen van de stage.

Het gemiddeld aantal uren per werkdag (per stage of leerbaan) wordt berekend als de verhouding tussen de omvang van de stage in uren (dit is het totaal aantal voor deze stage / leerbaan te werken uren) en de geplande duur van de stage (in werkdagen).

Vervolgens wordt het gemiddeld aantal uren per week bepaald als $5 * \text{het gemiddeld aantal uren per werkdag}$. Stages en leerbanen met een bijzonder hoog of bijzonder laag aantal uren per werkdag (minder dan 0,4 uur⁸ of meer dan 8,0 uur) blijven hierbij buiten beschouwing.

Ook het aantal weken dat een stage of leerbaan duurt wordt bepaald; hierbij wordt niet zozeer het totaal aantal weken geteld, maar het aantal weken *in het afgelopen schooljaar*. Dit is bijvoorbeeld relevant voor BBL leerbanen die een looptijd van meerdere jaren kunnen hebben.

Deze gegevens zijn voor beide type leerwegen afgeleid. Zowel voor BOL als voor BBL laten de gegevens een behoorlijke variatie zien, zowel qua aantal uren per week als aantal weken per jaar. In overleg met SBB is besloten om voor de BBL leerwegen uit te gaan van een vaste werkbelasting:

- elke leerbaan duurt precies 1 schooljaar; dit komt neer op 52 weken per jaar
- elke leerbaan telt evenveel uren per week; we gaan uit van 3 dagen / week (omdat de data laten zien dat BPVOs die voor BBL leerwegen zijn afgesloten gemiddeld ongeveer 3 dagen per week in beslag nemen).

Het resultaat van deze syntax is een apart databestand waarin per kwalificatie en per leerweg twee kenmerken van de omvang van stages en leerbanen zijn opgenomen:

- het gemiddeld aantal uren per week
- het gemiddeld aantal weken per jaar

Voor BBL leerbanen is dit voor alle kwalificaties hetzelfde, voor BOL stages kan dit variëren.

5.8 **_8_aanbod_jaar_Tplus1.sps**

Het doel van deze syntax is om het aanbod van stages en leerbanen voor het lopende schooljaar te bepalen. Dit bestand wordt op een aantal manieren aangepast en uitgebreid.

⁸ Dit komt overeen met minder dan 2 uur per week.



5.8.1 *Bepaal gemiddelde omvang huidige aanbod*

Voor de lopende BPVOs kan de omvang van het huidige aanbod bepaald worden op dezelfde manier als in syntax_7_ is gedaan; het enige verschil is dat nu gekeken wordt naar het deel van de lopende BPVOs dat in het lopende schooljaar valt (in syntax_7_ is gekeken naar de omvang van BPVOs in het voorafgaande schooljaar).

Voor de openstaande vacatures ontbreekt deze informatie; hiervoor wordt gebruik gemaakt van de correctiefactor uren_week die in syntax_4_ is afgeleid (op basis van informatie uit de leerbedrijvenquête).

Hierna wordt, per kwalificatie en leerweg, de gemiddelde omvang per stage en leerbaan bepaald. Dit is het product van het gemiddeld aantal uren per week en het gemiddeld aantal weken per jaar.

5.8.2 *Prognose van groei bij leerbedrijven die al in huidige aanbodbestand zitten*

In syntax_4_ is de correctiefactor inschrijvingen bepaald. Voor de bedrijven uit de leerbedrijvenquête die in het huidige aanbodbestand zitten, geeft deze factor de verhouding weer tussen het verwachte aanbod (aantal verwachte stages of leerbanen) en het huidige aanbod (aantal openstaande vacatures en lopende BPVOs).

Deze factor wordt nu (per kwalificatie, leerweg en regio) vermenigvuldigd met het totale huidige aanbod. Hierdoor ontstaat een prognose van het aanbod van stages en leerbanen bij leerbedrijven die al in het huidige aanbodbestand zitten.

5.8.3 *Prognose van groei bij leerbedrijven die nog niet in huidige aanbodbestand zitten*

De correctiefactor ontbreken leerbedrijven is ook in syntax_4_ bepaald, op basis van de bedrijven uit de leerbedrijvenquête die *niet* in het huidige aanbodbestand zitten. Deze correctiefactor is per sectorkamer en leerweg bepaald en geeft (in absolute cijfers) weer met hoeveel het huidige aanbod verhoogd (of verlaagd) moet worden.

Om deze correctie over alle kwalificaties, leerwegen en regio's te verdelen, wordt eerst per sectorkamer en leerweg de relatieve verdeling van het huidige aanbod over de onderliggende kwalificaties en regio's bepaald. Dit levert verschillende verhoudingen op. Door deze verhoudingen met de correctiefactor ontbrekende leerbedrijven te vermenigvuldigen ontstaat per kwalificatie, leerweg en regio een prognose van het aanbod voor het lopende schooljaar door leerbedrijven die nog niet in het huidige aanbodbestand zitten.

5.8.4 *Prognose totale aanbod op nationaal niveau*

De twee hiervoor beschreven correcties kunnen bij elkaar opgeteld worden. Dit levert een prognose van het totale aanbod van stages en leerbanen voor het lopende schooljaar van alle erkende leerbedrijven. Deze gegevens zijn per regio beschikbaar maar nog niet nationaal. Daarom worden de prognoses per kwalificatie en leerweg geaggregeerd om tot een prognose van het aanbod op nationaal niveau te komen.

5.9 ***_9_vraag_jaar_T.sps***

5.9.1 *Combineer bestanden met aantal inschrijvingen en gemiddeld aantal uren per BPVO*

De basis voor het huidige vraagbestand is in syntax_6_ gelegd. Hier is een bestand gemaakt waarin per kwalificatie en leerweg het totaal aantal ingeschreven studenten voor het afgelopen schooljaar is opgenomen.



Dit bestand wordt geopend, waarna informatie over de gemiddelde omvang van BPVOs (wat in `syntax_7_` is bepaald) wordt toegevoegd.

5.9.2 *De verhouding tussen aantal inschrijvingen en de vraag naar stages*

Het openstaande bestand bevat informatie over het aantal ingeschreven studenten in het afgelopen schooljaar. Dit wordt verder uitgebreid met informatie over het aantal afgesloten BPVOs in het afgelopen schooljaar (dit is in `syntax_5_` bepaald).

Per kwalificatie en leerweg is het nu mogelijk om het gemiddeld aantal afgesloten BPVOs per ingeschreven leerling te bepalen.

Voor de leerweg BOL wordt deze verhouding gebruikt om (per kwalificatie *en regio*) de vraag naar BPVOs te bepalen: de vraag naar BPVOs wordt berekend door het aantal ingeschreven studenten met deze verhouding te vermenigvuldigen. Het resultaat is vrijwel gelijk aan het aantal afgesloten BPVOs (immers, de verhouding is het aantal afgesloten BPVOs gedeeld door het aantal ingeschreven studenten).

Voor de leerweg BBL wordt de vraag naar BPVOs gelijkgesteld aan het aantal ingeschreven studenten. Dit is gedaan omdat een student zich niet voor een BBL opleiding kan inschrijven zonder een leerbaan bij een erkend leerbedrijf te hebben. Het is daarom opvallend dat gemiddeld genomen voor de BBL leerweg het aantal BPVOs per leerling 1,4 is en niet 1,0. Bij een volgende editie van de Kans op Stage zou dit verder onderzocht kunnen worden.

5.10 **`_10_vraag_jaar_Tplus1.sps`**

Het ministerie van OCW publiceert ramingen over het aantal studenten in het MBO, naar type leerweg en sector. Op basis van deze ramingen kan per leerweg en sector de groeivoet van het aantal studenten bepaald worden. We nemen aan dat het aantal BPVOs per leerling niet verandert. Hierdoor volstaat het om de vraag naar stages in leerbanen in het afgelopen schooljaar te vermenigvuldigen met de groeivoet van het aantal studenten om een prognose te maken van de vraag naar stages en leerbanen voor het huidige schooljaar.

5.11 **`_11_regionale_mobiliteit.sps`**

De regionale mobiliteit wordt bepaald op basis van het bestand met informatie over alle BPVOs uit het afgelopen schooljaar. Alleen waarnemingen over de leerwegen Bol en BBL, waarvoor de kwalificatie, woonregio en regio van leerbedrijf bekend zijn, kunnen hiervoor gebruikt worden.

5.11.1 *Bereken bruto en netto in- en uitstroom*

Per crebo en per leerweg worden de volgende kengetallen per regio berekend:

- `studenten_#`: aantal studenten dat in regio # woont (de noemer)
- `bruto_uitstroom_abs_regio_#`: aantal studenten dat in regio # woont maar elders werkt
- `bruto_instroom_abs_regio_#`: aantal studenten dat NIET in regio # woont maar er wel werkt

(`_abs` staat voor absoluut: het gaat hier om het absolute aantal studenten)

Hiermee kunnen in- en uitstroomcijfers berekend worden:

- $\text{Bruto_uitstroom_rel_regio \#} = \text{bruto_uitstroom_abs_regio_#} / \text{studenten_#}$
- $\text{Bruto_instroom_rel_regio \#} = \text{bruto_instroom_abs_regio_#} / \text{studenten_#}$
- $\text{Netto_instroom_rel_regio \#} = \text{bruto_instroom_rel_regio \#} - \text{bruto_uitstroom_rel_regio \#}$



(_rel staat voor relatief: het gaat hier om een verhouding)

Voordat deze kengetallen berekend worden, wordt de correctiefactor dossier gebruikt om rekening te kunnen houden met informatie die enkel op kwalificatiedossierniveau beschikbaar is.

5.11.2 *Bereken indicator 3 voor regionale mobiliteit*

Indicator drie moet (net als indicatoren 1 en 2) aan de volgende randvoorwaarden voldoen:

- Een score van 1,0 staat voor evenwicht.
- Een score lager dan 1,0 staat voor een relatief lage kans op een stage of leerbaan; in het geval van regionale mobiliteit wil dit zeggen dat er meer uitstroom dan instroom is.
- Een score hoger dan 1,0 staat voor een relatief hoge kans op een stage of leerbaan; in het geval van regionale mobiliteit wil dit zeggen dat er meer instroom dan uitstroom is.

Het uitgangspunt voor indicator 3 is de relatieve netto instroom. Deze variabele heeft de volgende kenmerken:

- netto_instroom_relatief ligt per definitie tussen -1 en oneindig;
- een waarde tussen -1 en 0 staat voor netto uitstroom;
- een waarde van 0 betekent dat er per saldo geen in- of uitstroom is;
- een positieve waarde staat voor netto uitstroom.

Indicator 3 wordt nu als volgt gedefinieerd:

- Indicator 3 = netto_instroom_relatief + 1;
- Als netto_instroom_relatief > 1 geldt (en dus indicator 3 > 2), dan wordt indicator 3 op 2 gesteld.

5.12 ***_12_indicatoren.sps***

In deze syntax worden de drie bestanden met de prognose van de vraag, prognose van het aanbod en de regionale dynamiek samengevoegd. Hierna kunnen de eerste drie indicatoren berekend worden en vervolgens de kans op stage.

5.12.1 *Combineer prognose aanbod met prognose vraag*

De volgende twee bestanden worden gecombineerd:

- bestand met prognose van het aanbod van stages en leerbanen (gecreëerd in syntax_8_);
- bestand met prognose van de vraag naar stages en leerbanen (gecreëerd in syntax_10_).

De bestanden worden gekoppeld op basis van crebocode, type leerweg en arbeidsmarktregio.

Vervolgens wordt gecontroleerd of voor alle waarnemingen in deze dataset het kwalificatienummer, kwalificatiedossier, marktsegment en sectorkamer bekend zijn. Voor sommige waarnemingen blijkt het marktsegment te ontbreken. Dit wordt alsnog bepaald (op basis van crebocode).

5.12.2 *Bepaal het aantal BPVOs in het lopende schooljaar*

Er is nog 1 variabele die ontbreekt om de indicatoren te kunnen berekenen: voor indicator 2 moet nog een prognose gemaakt worden van het aanbod in termen van aantal BPVOs. Dit wordt nu berekend, door het aantal BPVOs in het afgelopen schooljaar (bepaald in syntax_5_) te vermenigvuldigen met de groeivoet van BPVOs.



De groeivoet van het aantal BPVOs is in `syntax_4_` bepaald, op basis van de resultaten van de leerbedrijvenquête. Voor veel kwalificaties geldt dat deze groeivoet ontbreekt. Dit zijn vooral kwalificaties met relatief weinig studenten; deze kwalificaties komen maar weinig of niet in de leerbedrijvenquête voor, waardoor het niet mogelijk was om de groeivoet te bepalen. Voor deze kwalificaties zijn wel de groeivoeten op hoger aggregatieniveau beschikbaar (per kwalificatiedossier en leerweg, per marktsegment en leerweg, en per leerweg).

De groeivoet van het aantal BPVOs wordt daarom opnieuw berekend, op basis van dezelfde `syntax_4_`, maar op basis van de uitgebreidere dataset.

Ruimte voor verbetering

De berekening van de definitieve groeivoet van het aantal BPVOs in `syntax_4_` kan verwijderd worden; het gebeurt nu dubbel (in `syntax_4_` en in `syntax_12_`), wat voor verwarring kan zorgen en een kans op fouten introduceert.

5.12.3 *Berekening van indicatoren 1, 2 en 3*

Indicatoren 1 en 2 worden nu berekend. Indicator 1 wordt op basis van de volgende formule berekend:

- $Indicator_1 = \frac{aanbod_jaar_Tplus1_uren}{vraag_jaar_Tplus1_uren}$ (de prognose van het aanbod in het huidige schooljaar (jaar T plus 1), in uren; gedeeld door de prognose van de vraag in het huidige schooljaar, in uren).

Indicator 2 is in eerste instantie als volgt gedefinieerd:

- $Indicator_2_bpvos = \frac{BPVOs_jaar_Tplus1}{vraag_jaar_Tplus1}$ (de prognose van het aantal BPVOs dat in het huidige schooljaar afgesloten zal worden; gedeeld door de prognose van de vraag in het huidige schooljaar, in BPVOs)⁹.

Een risico van deze indicator is dat de omvang van de BPVOs aan de aanbodzijde niet overeen hoeft te komen met de omvang van de BPVOs aan de vraagzijde. Dat zou in een systematische over- of onderschatting van indicator 2 kunnen resulteren. Daarom is er een tweede versie van indicator 2 gedefinieerd, waarbij zowel aan de aanbod- als aan de vraagzijde het aantal BPVOs (per kwalificatie en leerweg) vermenigvuldigd wordt met de gemiddelde verwachte duur van de BPVOs:

- $Indicator_2_uren = \frac{BPVOs_jaar_Tplus1 * uren_jaar_aanbod}{vraag_jaar_Tplus1_uren}$.

Indicator 3 is al in `syntax_11_` berekend en wordt nu aan de dataset toegevoegd.

Ruimte voor verbetering

Het is verwarrend dat er twee versies van indicator 2 zijn. Voor de volgende keer moet onderzocht worden welke van deze twee de voorkeur geniet, waarna alleen deze berekend moet worden.

5.12.4 *Berekening van de kans op stage*

De kans op stage wordt in eerste instantie op landelijk niveau bepaald. De indicator voor de regionale mobiliteit speelt hier nog geen rol bij.

⁹ Voor de BOL leerweg geldt dat indicator_2_bpvos vrijwel gelijk is aan de groeivoet van het aantal BPVOs. Het aantal BPVOs in jaar t plus 1 wordt berekend door het aantal BPVOs in jaar T te vermenigvuldigen met de groeivoet van het aantal BPVOs. De vraag in jaar t plus 1 wordt berekend door de vraag in jaar t te vermenigvuldigen met de groeivoet van het aantal inschrijvingen. Omdat de groeivoet van het aantal inschrijvingen niet veel varieert (en slechts onderscheid naar 4 sectoren maakt) en de vraag in jaar t voor BOL min of meer gelijk is aan het aantal BPVOs in jaar t (zie par. 5.9.2), is de vraag in jaar t plus 1 (in BPVOs) sterk gecorreleerd met het aantal BPVOs in jaar t. Indicator 2 kan hierdoor geschreven worden als $(aantal\ BPVOs\ in\ jaar\ t) * (groeivoet\ BPVOs) / (aantal\ BPVOs\ in\ jaar\ t) = (groeivoet\ BPVOs)$. Dit blijkt te kloppen: voor de opleidingen waarvoor het aantal BPVOs in jaar t en de groeivoet van het aantal BPVOs bekend is, is de correlatie tussen indicator_2_bpvos en de groeivoet van het aantal BPVOs gelijk aan 0,995.



Op landelijk niveau zijn er vier indicatoren beschikbaar: indicator_1, indicator_2_bpvos, indicator_2_uren en indicator_4 (dit is de kans op stage van het vorig jaar). De eerste drie indicatoren zijn zo berekend dat ze aan de volgende voorwaarden voldoen:

- Ze zijn per definitie positief
- Een score van 1,0 staat voor evenwicht tussen vraag en aanbod.
- Een score lager dan 1,0 staat voor een relatief lage kans op een stage of leerbaan (de vraag naar stages is hoger dan het aanbod).
- Een score hoger dan 1,0 staat voor een relatief hoge kans op een stage of leerbaan (het aanbod van stages is hoger dan de vraag).

De vierde indicator kan zo worden herschaald dat ook aan deze eisen wordt voldaan.

De kans op stage wordt in eerste instantie berekend als een gewogen gemiddelde van de verschillende indicatoren. Dit gewogen gemiddelde zal dus ook een continue variabele zijn die aan bovenstaande kenmerken voldoet.

De variantie en spreiding van de verschillende indicatoren laat grote onderlinge verschillen zien. Als hier niet voor gecorrigeerd wordt, dan zal de spreiding in het gewogen gemiddelde vooral bepaald worden door de indicator met de hoogste spreiding. Dit is niet wenselijk, daarom worden de eerste drie indicatoren gestandaardiseerd.

De keuze voor de gewichten van de verschillende indicatoren is in overleg met SBB bepaald: voor verschillende keuzes zijn de kansen op stage bepaald, waarna de herkenbaarheid en plausibiliteit van de verschillende resultaten door twee experts van SBB vergeleken is.

Aan het begin van dit proces is besloten om geen gebruik te maken van de resultaten van het vorige jaar. Vervolgens is besloten om een onderscheid te maken tussen indicator_2_bpvos (die in aantallen BPVOs is gedefinieerd) en de andere twee indicatoren (die beide in uren gedefinieerd zijn). Om inzicht te krijgen in de invloed van verschillende gewichten is de volgende formule gebruikt:

$$\text{Kans_op_stage} = \text{gewicht} * \text{indicator_2_bpvos} + (1 - \text{gewicht}) * [\text{alfa} * \text{indicator_1} + (1 - \text{alfa}) * \text{indicator_2_uren}].$$

Na de uitkomsten voor verschillende gewichten vergeleken te hebben, is de volgende keuze gemaakt:

- gewicht = 0.75 (dit is het gewicht van indicator_2_BPVOs in de kans op stage);
- alfa = 1 (dit is het relatieve gewicht van indicator_1 ten opzichte van indicator_2_uren).

De keuze voor alfa=1 houdt in dat indicator_2_uren niet gebruikt is bij het bepalen van de kans op stage: op landelijk niveau is de kans op stage voor 75% bepaald door indicator_2_bpvo en voor 25% door indicator_1.

Op regionaal niveau is de kans op stage volgens de volgende formule afgeleid:

$$\text{Kans_op_stage_regionaal} = \text{gewicht_regionaal} * \text{kans_op_stage} + (1 - \text{gewicht_regionaal}) * \text{indicator_3}.$$

Met andere woorden: de regionale verschillen in de kans op stage worden enkel door de regionale mobiliteit bepaald.



Ook hier is de keuze voor gewicht_regionaal¹⁰ bepaald door de herkenbaarheid en plausibiliteit van de uitkomsten van verschillende gewichten te laten vergelijken door twee experts van SBB. Dit resulteerde in de volgende keuze:

- gewicht_regionaal = 0.8.

Zowel de nationale als de regionale variant van de kans op stage zijn nu nog continue variabelen: ze kunnen elke positieve waarde aannemen (al liggen ze vooral tussen 0 en 2). De laatste stap is om de kans op stage in te dikken in een variabele met vijf antwoordcategorieën:

- "gering": bij een waarde tussen 0 en 0,7;
- "matig": bij een waarde tussen 0,7 en 0,9;
- "voldoende": bij een waarde tussen 0,9 en 1,1;
- "ruim voldoende": bij een waarde tussen 1,1 en 1,3;
- "goed": bij een waarde van 1,3 of meer.

Ruimte voor onderzoek

Voor een volgende editie van Kans op Stage kunnen andere gewichten gekozen worden. Dit betekent dat de Kansen op stage niet goed direct met elkaar vergeleken kunnen worden. Als SBB ooit onderzoek zou willen (laten) doen naar de ontwikkelingen van de kans op stage, dan kan overwogen worden om de ontwikkelingen van de onderliggende indicatoren te onderzoeken.

¹⁰Gewicht_regionaal geeft aan in welke mate de regionale kans op stage bepaald wordt door de *landelijke* kans op stage.

