

ERASMUS UNIVERSITEIT ROTTERDAM

Erasmus School of Economics

Bachelorscriptie Economie en Bedrijfseconomie

Het effect van grootte van de opleiding op kans op werk op mbo niveau 4

Naam Student: Bas Stegehuis

Studentnummer: 371247

Begeleider: Matthijs Oosterveen

Datum definitieve versie: 6-1-2019

Het geschrevene in deze scriptie is de opvatting van de auteur en niet noodzakelijk die van Erasmus School of Economics of Erasmus Universiteit Rotterdam.

Abstract

Deze scriptie onderzoekt de relatie tussen de grootte van opleidingen op mbo niveau 4 en de kans op werk voor afgestudeerden. De verwachting is dat schaalvoordelen in het onderwijs de arbeidsmarktkansen voor studenten aan grote opleidingen vergroten. Tegelijkertijd kan het *class size-effect* ervoor zorgen dat juist kleine opleidingen het goed doen op de arbeidsmarkt. Er worden verschillende OLS-regressiemodellen gebruikt om het effect van grootte van de opleiding te testen. Uit deze modellen blijkt dat de grootte van de opleiding geen doorslaggevende factor is in de kans op werk na de opleiding. Om een compleet antwoord te geven op de hoofdvraag is meer onderzoek nodig, met betere data en/of andere methoden.

Inhoud

Inleiding	4
Theoretisch kader	6
Literatuur	6
Schaalvoordelen	6
Class-size effect	7
Hypothese	7
Data	8
DUO	8
CBS	8
Data samenvatting	9
Methodologie	11
De regressievergelijking	11
Percentage werk	11
Gaus-Markov-condities	11
Lineariteit	11
Random sampling of observations	11
Zero conditional mean	12
Multicollineariteit	13
Heteroscedasticiteit en autocorrelatie	14
Extreme waarden	14
Grootte van de opleiding	15
Resultaten	17
Beperkingen	20
Data	20
Methodologie	21
Conclusies	22
Referenties	23
Literatuur	23
Data	23
Appendix	24
Methodologie	24

Inleiding

Binnen het onderwijs speelt op bijna ieder niveau een discussie over de wenselijkheid van massaal onderwijs. Er zijn nog nooit zo veel hbo-studenten geweest als dit jaar, en ook het wo-onderwijs groeit hard (NU.nl, 2018). Ook in het mbo is dit een belangrijk vraagstuk. De afgelopen twintig jaar zijn roc's steeds groter geworden (Blank en Van Heezik, 2015). Het ligt hierbij voor de hand te vragen wat deze schaalvergroting in het onderwijs doet met de kwaliteit van het onderwijs. Er is behoorlijk wat onderzoek gedaan naar het effect van de grootte van een klas op de kwaliteit van het onderwijs. Er is echter nog weinig onderzoek naar het effect van de grootte van de opleiding op de kwaliteit van het onderwijs, en specifiek op de arbeidsmarktkansen van afgestudeerden.

Dit onderzoek richt zich op een statistisch onderbouwd antwoord op die vraag. Dat is nodig, omdat er de komende jaren belangrijke besluiten genomen moeten worden over schaalvergroting in het mbo. Zo zei de Minister van Onderwijs, Cultuur en Wetenschappen in 2015 in de toelichting op de Beleidsregel Macrodoelmatigheid Beroepsonderwijs:

“Voorts beoogt de wet een doelmatigere spreiding van het opleidingsaanbod. Een aanzienlijk deel van de beroepsopleidingen kent een relatief gering aantal studenten. Een doelmatigere spreiding maakt het in stand houden van dure maar waardevolle beroepsopleidingen beter mogelijk en leidt tot een doelmatigere besteding van publieke middelen, wat ten goede komt aan de kwaliteit en continuïteit van het opleidingsaanbod.”

Blijkbaar zijn kleine opleiding duur om in stand te houden, maar vindt de Minister het toch waardevol om ze te behouden. In de praktijk is al zichtbaar dat roc's worstelen met de vraag hoe ze hiermee om moeten gaan. Er wordt bijvoorbeeld gekeken naar het samenvoegen van kleine opleidingen om ze in stand te kunnen houden (Gelderblom et al, 2016).

Binnen het mbo bestaat er een grote spreiding van het aantal deelnemers over opleidingen, zoals later ook aangetoond zal worden. Het is van belang voor onderwijsinstellingen om te weten of de grootte van een opleiding effect heeft op de arbeidsmarktkansen van de studenten, als zij daadwerkelijk overwegen kleine opleidingen stop te zetten of samen te voegen. De hoofdvraag van deze scriptie luidt dan ook:

“Wat is de invloed van de grootte van een mbo4-opleiding op de kans op een baan voor een student?”

Vanwege de beperkte tijd voor dit onderzoek en omdat er grote verschillen zijn tussen de opleidingen en de studenten op verschillende niveaus gaat dit onderzoek specifiek over studenten op het hoogste niveau in het mbo, niveau 4.

In het volgende hoofdstuk wordt gekeken naar twee verschillende effecten die kunnen optreden bij schaalvergroting: schaalvoordelen en het *class size*-effect. Deze effecten worden besproken en aan de hand van de beschikbare literatuur daarover wordt er een hypothese opgesteld om de hoofdvraag te toetsen. Daarna worden de twee gebruikte datasets toegelicht en wordt duidelijk hoe deze zijn samengevoegd. In het hoofdstuk *methodologie* komen de voorwaarden voor een *OLS*-regressie aan bod en hoe de dataset hierop getest is. Daarnaast worden verschillende regressievergelijkingen opgesteld. De uitkomsten hiervan zijn in het hoofdstuk *Resultaten* te vinden. In dat hoofdstuk wordt ook de geformuleerde hypothese getest. In de laatste twee hoofdstukken worden de beperkingen van dit onderzoek besproken en wordt voorzichtig een antwoord geformuleerd op de hoofdvraag.

Theoretisch kader

Literatuur

De grootte van een opleiding kan op verschillende manieren effect hebben op de kans op werk na de opleiding. Hieronder worden twee manieren besproken aan de hand van de literatuur.

Schaalvoordelen

In het onderwijs kan net als in andere sectoren sprake zijn van schaaffecten. Die treden op als de gemiddelde kosten per student dalen naarmate het aantal studenten toeneemt. De financiering in het middelbaar beroepsonderwijs is grotendeels gebaseerd op het aantal studenten van een instelling. Als schaalvoordelen optreden, kunnen grotere opleidingen hun studenten meer bieden dan kleine opleidingen. Omdat de instellingen het geld zelf over de opleidingen kunnen verdelen, hoeven deze schaalvoordelen niet meteen te leiden tot beter onderwijs voor grotere opleidingen. Instellingen kunnen kleine, dure opleidingen die zij belangrijk vinden overeind houden door ze te financieren met geld dat ze ‘overhouden’ aan de grotere opleidingen (Gelderblom et al, 2017).

Er zijn sterke aanwijzingen dat er financiële schaalvoordelen in het onderwijs zijn (Blank en Van Heezik, 2015). De afgelopen decennia zijn de Nederlandse ROC-instellingen sterk gegroeid. Zo is tussen 1980 en 2012 de gemiddelde omzet per ROC gestegen van 6 miljoen naar 66 miljoen. In de jaren 1980 – 2000 kwam de schaalvergroting vooral door fusies, in de jaren daarna door het toenemende aantal leerlingen (Blank en Van Heezik, 2015).

Schaalvoordelen kunnen ook niet-financieel van aard zijn. Specifiek voor de kans op werk is het denkbaar dat schaalvoordelen optreden doordat het voor werkgevers makkelijker is om invloed op de opleiding uit te oefenen als zij te maken hebben met één grote opleiding in plaats van meerdere kleine opleidingen. Als werkgevers een belangrijkere stem hebben in het onderwijs, ligt het voor de hand dat ze daarna sneller afgestudeerden aan zullen nemen. Zij hebben op de opleiding immers de vaardigheden geleerd die voor de toekomstige werkgever van belang zijn. Dit gebeurt zeker in het mbo, voornamelijk op de lagere niveaus. Tegelijkertijd zijn instellingen zich ervan bewust dat een opleiding meer is dan het voorbereiden op de arbeidsmarkt. Studenten krijgen ook volgen ook vakken als burgerschap, Nederlands en Engels (Gelderblom et al, 2016).

Het lijkt gezien het bovenstaande aannemelijk dat schaalvoordelen in het mbo kunnen leiden tot een grotere kans op werk voor afgestudeerden. Dit leidt tot de verwachting dat de grootte van de opleiding de kans op werk positief beïnvloedt.

Class-size effect

Bij klassikaal onderwijs kan de verhouding studenten per docent de kwaliteit van de opleiding beïnvloeden. Omdat er bij (zeer) kleine opleidingen bijna per definitie ook kleine klassen horen, zou dit een positieve invloed kunnen hebben op de kans op werk. Tegelijkertijd betekent een grote opleiding niet meteen dat het aantal studenten per klas groot is. Ook bij grote opleidingen is kleinschalig onderwijs mogelijk.

Een meta-studie naar verschillende Amerikaanse studies naar het *class size-effect* laat zien dat onderwijsresultaten en kans op goed betaald werk na de opleiding kunnen verbeteren door klassen te verkleinen, maar dat de verbetering die optreedt relatief klein is (Chingos, 2012). In dit onderzoek worden negen eerdere onderzoeken tussen 1999 en 2012 besproken, waaronder bijvoorbeeld dat van Krueger (1999) naar het effect van het verkleinen van klassen op test-scores. In Hongkong werd in 2015 een studie uitgevoerd waarbij leraren grote (± 40) en kleine (± 25) klassen Engels doceerden en onderzoekers in de lessen observeerden. De relatie tussen studenten onderling en tussen studenten en docenten was beter in de kleinere klassen. Studenten in kleine klassen deden actiever mee in de les en zij voelden zich gelukkiger (Harfitt en Tsui, 2015). Of dit ook leidt tot een grotere kans op werk is niet duidelijk.

De verwachting is gezien het bovenstaande dat het class-size effect zorgt voor een negatieve invloed van de grootte van de opleiding op de kans op werk.

Hypothese

De theorie hierboven wijst naar twee mechanismen die via schaalgrootte effect kunnen hebben op de kans op werk. De eerste is de aanwezigheid van schaalvoordelen, wat een positief effect kan hebben op de kans op werk. Het tweede is het class-size effect, wat een negatief effect kan hebben op de kans op werk. Het is niet duidelijk hoe groot het effect van deze twee mechanismen is op de kans op werk of welk effect de overhand heeft. Theoretisch is het dus onduidelijk wat het effect is van grootte van de opleiding op kans op werk. De hoofdvraag van deze scriptie luidt daarom:

“Wat is het effect van de grootte van de opleiding op de kans op werk na het afstuderen op mbo niveau 4?”

De hypothese die daarbij hoort luidt als volgt:

H₁: Grootte van de opleiding op mbo4 heeft een significant effect op de kans op werk na de opleiding

Data

Voor dit onderzoek zijn twee databestanden gebruikt en bewerkt. Hieronder worden de datasets toegelicht.

DUO

Het eerste bestand dat is gebruikt bevat een overzicht van alle opleidingen op niveau 4 per instelling van alle roc's in Nederland. In dit bestand zijn de aantallen deelnemers per opleiding over de jaren 2013 tot en met 2017 te vinden. Iedere opleiding heeft een unieke *crebo-code*, waardoor opleidingen van verschillende instellingen met elkaar te vergelijken zijn. De cijfers over aantallen zijn afkomstig van het Basisregister Onderwijs, dat onder DUO valt. Alle studenten die in Nederland een opleiding volgen staan hierin geregistreerd.

CBS

Het CBS heeft een bestand beschikbaar gesteld met gegevens over schoolverlaters van alle opleidingen in Nederland per instelling op crebo-code. Daarin is het werkloosheidspercentage te vinden van iedere individuele instelling en het werkloosheidspercentage per opleiding van alle instellingen, zowel totaal als uitgesplitst naar bol en bbl. De data zijn beschikbaar voor een half jaar en anderhalf jaar na het afstuderen. Voor dit onderzoek worden studenten onderzocht die zijn afgestudeerd aan het eind van schooljaar 2014/15, dus de gegevens met betrekking tot werk gaan over eind 2015 en eind 2016. Voor de gegevens met betrekking tot werk heeft het CBS gebruik gemaakt van de registratie van het UWV. Het gaat uitsluitend om schoolverlaters, dus studenten die een vervolgopleiding zijn gaan doen worden hierin niet meegenomen. Opleidingen met minder dan vijf schoolverlaters zijn niet in deze dataset meegenomen.

De aantallen per opleiding in het CBS-bestand wijken sterk af van de duo-gegevens. DUO kijkt naar het aantal ingeschreven studenten op mbo4 in 2014 (dus over alle leerjaren), terwijl het CBS kijkt naar aantallen afgestudeerde schoolverlaters (van één leerjaar) met een diploma. Studenten die zijn afgehaakt of een vervolgopleiding zijn gaan doen worden in dit bestand buiten beschouwing gelaten. Van alle mbo-studenten stroomt ongeveer 35% door naar het hbo, maar dit verschilt sterk per opleiding (CBS, 2018). Dit kan een bias in de werkloosheidspercentages opleveren, omdat het waarschijnlijk de beter presterende studenten zijn die doorstromen, en die dus niet in de schoolverlaterscijfers worden meegenomen. De studenten die minder goed presteren, en dus niet kunnen doorstromen, hebben na de opleiding geen andere keus dan de arbeidsmarkt op te gaan. De arbeidsmarktkansen voor die groep zijn (los van het effect van de vervolgopleiding) waarschijnlijk ook lager dan die voor de groep die de mogelijkheid heeft om door te stromen. Hierdoor kan het

gebeuren dat het werkloosheidspercentage van het CBS hoger ligt dan het daadwerkelijke percentage.

In het hoofdstuk ‘beperkingen’ wordt stilgestaan bij de mogelijke verbeteringen om dit probleem op te lossen. Het werkloosheidspercentage wordt vastgesteld aan de hand van de CBS-definitie van voor 2015, dus iedereen die minder dan 12 uur in de week werkt wordt als werkloos geteld.

Data samenvatting

In deze thesis gaat het steeds om opleidingen per instelling. Iedere opleiding wordt geïdentificeerd aan de hand van een combinatie van drie factoren. De eerste factor is de BRIN-code van de instelling die de opleiding verzorgt. De tweede is de naam van de opleiding (deze zijn landelijk geregistreerd via *crebo*-codes en daardoor met elkaar vergelijkbaar tussen instellingen). De laatste factor is het type opleiding, bol of bbl. Een voorbeeld van een opleiding is dus: 00GT 10426 BBL. Dit is de bbl-variant van de opleiding verpleegkunde aan het ROC Albeda.

Binnen het mbo niveau 4 zijn er grote verschillen in de grootte van opleidingen. Er zijn relatief veel kleine opleidingen, negentig procent van alle opleidingen heeft minder dan 151 studenten. Er zijn vijf opleidingen met meer dan 1.000 studenten, de grootste opleiding telt er bijna 1.500. In figuur 1 is de verdeling van de aantallen studenten per opleiding te zien. In figuur 1 staan alle opleidingen die kleiner zijn dan 300 leerlingen, dat gaat om 97.3% van alle opleidingen. Voor de leesbaarheid zijn de overige opleidingen weggelaten.

In tabel 1 en tabel 2 zijn alle gebruikte data kort samengevat. In totaal zijn 3909 opleidingen geanalyseerd.

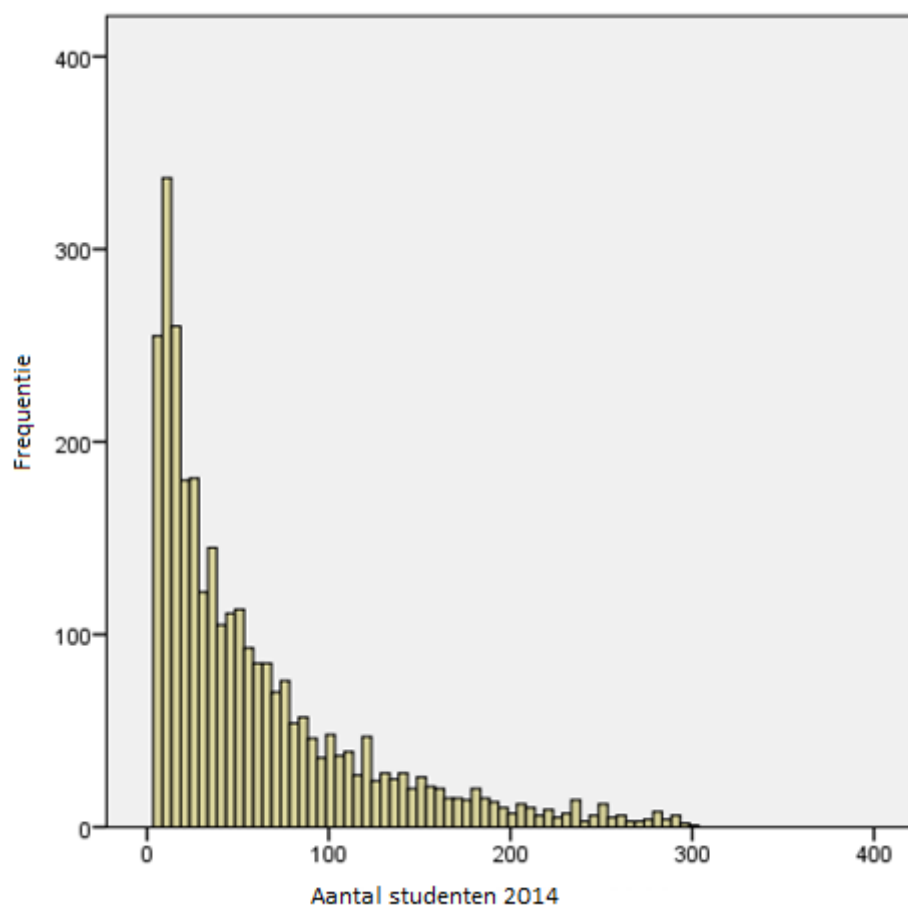
Tabel 1: Aantal studenten per opleiding en percentage werk per opleiding

Totaal aantal opleidingen	3909	
	Aantal studenten	Percentage werk
Gemiddelde	61	68,1%
Standaardafwijking	97	28,5%

Tabel 2: Kenmerken opleidingen

Kenmerken opleidingen		
Type BBL	27,6%	1079
Type BOL	72,4%	2830
Sector Economie	35,9%	1403
Sector Zorg	25,0%	977
Sector Techniek	33,1%	1295

Sector Groen	6,0%	234
Randstad	28,1%	1097



Figuur 1: Verdeling van het aantal studenten per opleiding (alleen opleidingen < 300, 97.3% van alle opleidingen)

Methodologie

De regressievergelijking

Om het effect van opleidingsgrootte op de kans op werk te meten wordt een *Ordinary Least Squares* regressie (OLS) gebruikt. Met die regressie wordt een model gemaakt om het percentage werk te schatten aan de hand van de grootte van de opleiding. Het doel hiervan is om de coëfficiënt voor grootte van de opleiding zo causaal mogelijk te schatten. Die vergelijking ziet er als volgt uit:

$$(1) \text{Percentage werk}_{ij} = \alpha_0 + \alpha_1 N_{ij} + \varepsilon_{ij}$$

Hierin staat N voor het aantal studenten in de opleiding, i voor opleiding i en j voor instelling j.

Percentage werk

Zoals eerder genoemd is het percentage werk op twee momenten gemeten. Voor studenten die zijn afgestudeerd in het jaar 2014/15 is dit zowel in oktober 2015 als in oktober 2016 gemeten. Alle regressies worden zowel uitgevoerd met het percentage werk in 2015 als in 2016.

Gaus-Markov-condities

Voordat een OLS regressie kan worden uitgevoerd, moet de data aan de *Gauss-Markov condities* voldoen. Deze condities worden hieronder toegelicht en waar nodig getest.

Lineariteit

De eerste voorwaarde van de Gaus-Markov condities is dat de afhankelijke variabele een lineaire functie is van de onafhankelijke variabelen en de error term.

Random sampling of observations

De tweede conditie houdt in dat steekproeven willekeurig getrokken moeten worden uit de populatie, dat het aantal observaties groter is dan het aantal onafhankelijke variabelen en dat de afhankelijke variabele de onafhankelijke variabelen niet beïnvloedt. In dat laatste geval is er een correlatie tussen de afhankelijke en onafhankelijke variabele, maar is een (mogelijke) causale relatie met OLS niet te vinden.

Hoewel het mogelijk is dat het percentage studenten dat werk vindt na de opleiding van invloed kan zijn op het aantal studenten dat kiest voor de opleiding, blijkt dat in de praktijk weinig het geval te zijn (De Koning et al, 2010). Als dat effect er al is, bijvoorbeeld doordat studenten ontmoedigd worden om een opleiding te kiezen vanwege de hoge werkloosheid van oud-studenten, werkt dat altijd met vertraging door. Het werkloosheidspercentage in 2015 kan nooit het aantal studenten in 2014 beïnvloeden.

Zero conditional mean

De derde conditie stelt dat de verwachte waarde voor de error term voor iedere willekeurige waarde van X gelijk moet zijn aan nul. Als dit niet het geval is, zit er een bias in de schatting van het model. Dit kan bijvoorbeeld komen doordat er een *omitted variable* is die via een onafhankelijke variabele de afhankelijke variabele beïnvloedt. Om dit tegen te gaan wordt een aantal controlevariabelen geïntroduceerd.

Type opleiding

Binnen het mbo zijn er twee typen opleidingen te onderscheiden: de beroepsopleidende leerweg (bol) en de beroepsbegeleidende leerweg (bbl). Een bol-student meldt zich aan bij een mbo-instelling en zal vervolgens een groot deel van de opleiding binnen die instelling doorbrengen, zoals dat ook gebruikelijk is bij een hbo- of academische opleiding. Naast de klassikale lessen zijn er ook stages binnen de bol-opleiding.

De bbl-opleiding werkt anders, en dat is voor dit onderzoek van wezenlijk belang. Een student die een bbl-opleiding wil gaan volgen, moet eerst zelf een werkgever zoeken. De werkgever moet een gecertificeerde opleider zijn. De student gaat vervolgens vier dagen in de week aan het werk bij zijn werkgever en brengt één dag in de week op de mbo-instelling door om algemene vakken te volgen.

Bbl-studenten vinden na de opleiding over het algemeen makkelijker een baan. In 2016 lag het percentage werkloosheid voor bol-studenten op mbo4 op 10%, terwijl dat voor bbl-studenten op mbo4 2% was (Meng en Sijbers, 2017). Er is een selectie-effect, alleen leerlingen die een baan kunnen vinden volgen een bbl-opleiding. Daarnaast hebben bbl-studenten aantoonbare werkervaring aan het eind van de opleiding. Vanwege dit verschil is het belangrijk om tijdens het onderzoek duidelijk verschil te maken tussen beide groepen studenten. Daarom wordt een dummy-variabele voor het type opleiding toegevoegd. De variabele neemt de waarde 0 aan voor bol-opleidingen en de waarde 1 voor bbl-opleidingen.

Regio

Er zijn binnen Nederland grote verschillen in werkloosheid tussen regio's. Zo is de werkloosheid in de Nederlandse grote steden hoger dan daarbuiten (CBS, 2018). Tegelijkertijd is er in de Randstad meer bedrijvigheid en economische productie dan daarbuiten. Daarom is het niet ondenkbaar dat de kans op een baan binnen de Randstad anders is dan de kans op een baan daarbuiten. Het CBS gebruikt als definitie van de Randstad 'een min of meer aaneengesloten gebied in het westen van het land waar de bevolkingsdichtheid het hoogst is'. Omdat de bevolkingsdichtheid hoger is, zijn er waarschijnlijk meer kansen voor schaalvergroting in de Randstad. Studenten in de Randstad hebben dus waarschijnlijk een grotere kans om deel te nemen aan een relatief grote opleiding, en ze hebben

waarschijnlijk een iets grotere baankans. Als er geen variabele voor Randstad wordt gebruikt, kan dit leiden tot het overschatten van het effect van grootte van de opleiding op baankans. Daarin wordt voor de Randstad-regio een dummy-variabele toegevoegd, die de waarde 0 aanneemt voor alle gebieden buiten de Randstad en 1 voor alle gebieden in de Randstad¹.

Sector

Het ligt voor de hand dat het uitmaakt voor baankansen welke specifieke opleiding er gevolgd wordt. In verschillende sectoren van de economie is de kans op een baan nu eenmaal groter dan in andere sectoren (Meng en Sijbers, 2017). Tegelijkertijd is de sector ook van belang voor de grootte van de opleiding, omdat bekend is dat bijvoorbeeld technische opleidingen over het algemeen specialistischer en dus kleinschaliger zijn dan economische opleidingen (Gelderblom et al, 2016). In dit onderzoek worden alle opleidingen ingedeeld in vier sectoren: groen, techniek, economie en gezondheidszorg. Binnen deze sectoren is er nog een grote diversiteit aan opleidingen. Zo vallen zowel de opleiding fijnmechanische techniek als de opleiding mediavormgeving onder de sector techniek. Het is mogelijk om meer specifieke indelingen te maken, maar dat maakt het weer lastiger om meer algemene conclusies te trekken. Er is sprake van een trade-off, waarbij vier sectoren een verdedigbare middenweg lijkt. Voor de sectoren wordt een dummy-variabele gebruikt.

Als de dummy-variabelen voor type opleiding, randstad en sector toegevoegd worden, ziet de vergelijking er als volgt uit:

$$(2) \text{Percentage werk}_{ij} = \alpha_0 + \alpha_1 N_{ij} + \alpha_2 bbl_{ij} + \alpha_3 randstad_{ij} + \alpha_4 economie_{ij} + \alpha_5 groen_{ij} + \alpha_6 techniek_{ij} + \varepsilon_{ij}$$

Multicollineariteit

De vierde conditie is het ontbreken van multicollineariteit. Als er sprake is van multicollineariteit, zijn er meerdere onafhankelijke variabelen die hetzelfde effect meten. Dit zou bijvoorbeeld kunnen spelen bij type opleiding en sector, omdat in de technische sector veel meer gebruik wordt gemaakt van bbl-opleidingen dan in de economische sector (Gelderblom et al, 2016). De aanwezigheid van multicollineariteit wordt gemeten aan de hand van de *tolerance factor* en de *VI factor* van de individuele onafhankelijke variabelen.

¹ De volgende door DUO aangehouden onderwijsregio's worden in dit onderzoek tot de Randstad gerekend: Utrecht en omstreken, Rijnmond groot, Zuid-Noord-Holland en Noord-Zuid-Holland

Er blijkt in de data geen sprake te zijn van significante multicollineariteit. De tolerance en VI-factoren geven geen reden tot zorg². De volledige tabellen zijn te vinden in de bijlage.

Heteroscedasticiteit en autocorrelatie

De laatste voorwaarde van Gaus-Markov is dat er geen sprake is van heteroscedasticiteit en autocorrelatie. Bij homoscedasticiteit zijn de waarden van de residuen niet afhankelijk van de waarde van de afhankelijke variabele. Als dit wel het geval is, is er sprake van heteroscedasticiteit. Omdat OLS een manier is om de lijn te vinden die de residuen minimaliseert, zullen grotere residuen een sterke invloed hebben op de regressielijn. Verder leidt heteroscedasticiteit tot *biased* standaardfouten. Dit kan leiden tot onjuiste conclusies met betrekking tot de significantie van de coëfficiënten van het model. Om te testen of er sprake is van heteroscedasticiteit wordt een Koenker test voor heteroscedasticiteit gebruikt. Hoewel net niet significant, geeft de Koenker test reden om te twifelen over heteroscedasticiteit. Daarom is voor de zekerheid ook een Breusch-Pagan test uitgevoerd, die wel wijst op heteroscedasticiteit. Om hiervoor te corrigeren worden robuuste standaard errors gebruikt.

Autocorrelatie treedt op als de error term van verschillende observaties met elkaar gecorreleerd zijn. Om dit te testen wordt bij iedere regressie een Durbin-Watson test gebruikt om autocorrelatie te detecteren. Uit de tests blijkt dat er een risico is dat de error termen positief gecorreleerd zijn. Hoewel autocorrelatie vooral bij tijdreeksen problemen opleveren, is het ook bij cross-sectie data belangrijk om na te gaan. Het kan bijvoorbeeld voorkomen als het type opleiding van student A de kans op werk voor student B beïnvloedt. Daarnaast kan het een aanwijzing zijn dat er een *omitted variable* is. Om mogelijke autocorrelatie op te lossen zijn de eerder genoemde robuuste standaard errors gebruikt. De genoemde tests zijn terug te vinden in de bijlage.

Extreme waarden

Voor de regressie uitgevoerd wordt, is het verstandig om de data te controleren op extreme waarden, omdat die een grote invloed kunnen hebben op de regressielijn. Omdat de afhankelijke variabele in dit onderzoek een percentage is, zijn de hoogst en laagst mogelijke waarde per definitie honderd en nul. Beide waarden komen geregeld voor in de dataset, allebei bij ongeveer 10% van de opleidingen. Dit gaat zowel om grote als om kleine opleidingen qua totaal aantal studenten, maar het zijn altijd opleidingen met een klein aantal afgestudeerden dat in 2014 de arbeidsmarkt op ging. Dit kan zijn omdat het een nieuwe opleiding is die in korte tijd groot is geworden, of omdat het een opleiding is waarbij het overgrote deel van de afgestudeerden doorstroomt naar het hbo.

² Voor de Tolerance Factor is <0,20 als grenswaarde gekozen, voor de VI-Factor >5

OLS houdt geen rekening met de grens van nul en honderd, wat betekent dat het uiteindelijke regressiemodel voor een opleiding een percentage boven de honderd zou kunnen voorspellen. Het doel van dit onderzoek is echter niet het bouwen van een model dat voorspellingen doet, maar het zo causaal mogelijk schatten van het effect van opleidingsgrootte. Daarom is toch voor een lineaire regressie gekozen.

Grootte van de opleiding

Het voorspellen van de kans op werkloosheid aan de hand van grootte van de opleiding levert een probleem op. In de dataset zitten veel zeer kleine opleidingen. Ongeveer een kwart van de opleidingen heeft minder dan 10 leerlingen. Hier zijn verschillende verklaringen voor. Dit kan gaan om opleidingen op maat, waarbij studenten min of meer dezelfde opleiding volgen met verschillende accenten. Deze opleidingen worden dan geregistreerd als verschillende individuele opleidingen. Daarnaast zijn er veel kleine bbl-opleidingen, wat voor de instelling geen probleem hoeft te zijn. De student volgt bij de werkgever de vakspecifieke opleiding, de algemene vakken worden aan de onderwijsinstelling gevolgd. Dit maakt het voor instellingen mogelijk om studenten van verschillende kleine opleidingen bij elkaar te zetten. Een derde mogelijkheid voor een zeer kleine opleiding is dat het gaat om een opleiding die wordt stopgezet, maar er nog een kleine groep is die nog niet is afgestudeerd.

Het probleem van deze groep is dat een kleine wijziging (een of twee studenten die geen werk vinden) een groot effect kan hebben op het percentage dat werk vindt na de opleiding. Om toch te kijken naar het effect van zeer kleine opleidingen is er een dummy-variabele aangemaakt voor zeer kleine opleidingen. Dit zijn alle opleidingen met minder dan 12 studenten. De variabele heeft de waarde nul voor alle opleidingen groter dan 11. De grens van 12 is gekozen omdat hier de cumulatieve grens van vijftientwintig procent ligt voor alle opleidingen. Op deze manier worden alle kleine opleidingen samengevoegd om het effect van de opleidingsgrootte te kunnen meten. De vergelijkingen zien er dan als volgt uit:

$$(3) \text{ Percentage werk}_{ij} = \alpha_0 + \alpha_1 \text{kleine opleiding}_{ij} + \varepsilon_{ij}$$

$$(4) \text{ Percentage werk}_{ij} = \alpha_0 + \alpha_1 \text{kleine opleiding}_{ij} + \alpha_2 \text{bbl}_{ij} + \alpha_3 \text{randstad}_{ij} + \alpha_4 \text{groen}_{ij} + \alpha_5 \text{techniek}_{ij} + \alpha_6 \text{economie}_{ij} + \varepsilon_{ij}$$

Om de invloed te meten van de grootste opleidingen, is er ook een dummy aangemaakt voor alle opleidingen groter dan 55, de grootste 10% van alle opleidingen. Als schaalvoordelen niet lineair zijn

maar slechts optreden voor de grootste opleidingen, zou deze variabele een significant positief effect moeten hebben op het percentage werkenden. De vergelijkingen zien er dan als volgt uit:

$$(5) \text{ Percentage werk}_{ij} = \alpha_0 + \alpha_1 \text{ grote opleiding}_{ij} + \varepsilon_{ij}$$

$$(6) \text{ Percentage werk}_{ij} = \alpha_0 + \alpha_1 \text{ grote opleiding}_{ij} + \alpha_2 \text{ bbl}_{ij} + \alpha_3 \text{ randstad}_{ij} + \alpha_4 \text{ groen}_{ij} + \alpha_5 \text{ techniek}_{ij} + \alpha_6 \text{ economie}_{ij} + \varepsilon_{ij}$$

Resultaten

Hieronder zullen de resultaten van de hierboven besproken regressies besproken en geïnterpreteerd worden. Onderaan dit hoofdstuk, in tabel 3, zijn de resultaten van alle regressies te vinden.

In vergelijking (1) heeft het aantal studenten aan een opleiding een statistisch significante invloed op de baankansen na de studie. Wel moet worden opgemerkt dat de grootte van de opleiding slechts een zeer klein deel van de variantie in baankansen verklaart. De verwachting is dat dit verandert als het aantal relevante variabelen toeneemt. Dat is het geval in vergelijking (2).

In vergelijking (2) zijn alle onafhankelijke variabelen behalve *Aantal studenten* dummy-variabelen. Zoals verwacht hebben de sector en het type opleiding een redelijk sterke invloed op de kans op een baan. De invloed van de variabele 'Randstad' is blijkbaar niet significant. Als voor de controlevariabelen wordt gecorrigeerd, valt echter het effect van grootte van de opleiding bijna helemaal weg. Grootte van de opleiding is in dit model niet meer statistisch of economisch significant. Daarnaast valt op dat de *adjusted R-squared* hoger is dan in vergelijking (1), maar dat deze nog steeds zeer laag is. Blijkbaar zijn er nog andere factoren die een sterke invloed hebben op het percentage werkenden na een opleiding. Te denken valt bijvoorbeeld aan de kwaliteit en beschikbaarheid van stageplekken tijdens de opleiding.

Ook als alleen gekeken wordt naar of een opleiding wel of niet klein is, zoals in vergelijking (3) en (4), valt op dat het effect van de grootte van de opleiding niet statistisch significant is.

Bij vergelijking (5) en (6) komt het patroon terug dat eerder ook te zien was. Als alleen de grootte van de opleiding als verklarende factor genomen wordt, is het een significante factor, hoewel de verklarende kracht van het model zeer laag is. Zodra de controlevariabelen worden toegevoegd, is de grootte van de opleiding niet meer significant.

Hoewel deze scriptie een aantal beperkingen kent die het lastig maken om sterke conclusies te trekken, moet aan de hand van de verkregen resultaten de hypothese H1 verworpen worden:

H₁: Grootte van de opleiding op mbo4 heeft een significant effect op de kans op werk na de opleiding

Behalve in vergelijking (1) en (5), die duidelijk incompleet zijn, is de variabele grootte van de opleiding in geen van de modellen statistisch significant. Het is aan de hand van deze scriptie niet te zeggen of de eerder besproken schaalvoordelen met betrekking tot baankansen en/of *class size-effects* aanwezig zijn in het mbo op niveau 4. Wel is duidelijk dat geen van deze effecten zo sterk is dat het het andere effect domineert en zichtbaar is in het effect van grootte van de opleiding op de kans op werk.

Tabel 3: Resultaten van de regressievergelijkingen

	(1)		(2)		(3)	
	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
Aantal studenten	0.10	6.11***	0.00	1.12		
Type BBL			2.61	2.52**		
Randstad			-1.01	-0.92		
Sector Economie			5.35	4.36***		
Sector Groen			3.07	1.38		
Sector Techniek			5.38	4.01***		
Kleine opleiding					-0.26	-0.27
Grote opleiding						
Adjusted R-squared	0.0050		0.0074		0.0002	
Aantal observaties	3909		3909		3909	

De significantie wordt weergegeven door asterisken: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tabel 3: Vervolg

Vergelijking	(4)		(5)		(6)	
	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
Aantal studenten						
Type BBL	2.72	2.58***			2.08	1.98**
Randstad	-1.01	-0.92			-1.01	-0.92
Sector Economie	5.38	4.39***			5.11	04.16***
Sector Groen	3.12	1.40			2.76	1.23
Sector Techniek	5.39	4.11***			5.11	3.87***
Kleine opleiding	1.09	1.06				
Grote opleiding			2.15	2.22**	0.89	0.87
Adjusted R-squared	0.0074		0.0010		0.0089	
Aantal observaties	3909		3909		3909	

De significantie wordt weergegeven door asterisken: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Beperkingen

Zoals eerder benoemd in het hoofdstuk *data* kent dit onderzoek een aantal beperkingen dat het gevaarlijk maakt om sterke conclusies te trekken. Dit hoofdstuk gaat daar verder op in en geeft een aantal suggesties om dit te verbeteren. De problemen kunnen worden opgesplitst in problemen met de data en beperkingen van de methodologie.

Data

Een van de grootste obstakels in deze scriptie is het definiëren van grootte van de opleiding. Het is voor de hand liggend om te kijken naar het aantal studenten dat bij een instelling staat ingeschreven, maar dat blijkt in de praktijk vaak af te wijken van de situatie in de klas, bijvoorbeeld doordat verschillende vergelijkbare kleine opleidingen gezamenlijk les krijgen van dezelfde docenten. Daarom is het bij een mogelijk vervolgonderzoek van belang om ook mensen te betrekken die ervaring hebben met het onderwijs, door bijvoorbeeld een kwalitatief deel aan het onderzoek toe te voegen. Dat kan bijvoorbeeld bestaan uit interviews met docenten of decanen.

In deze scriptie zijn data per opleiding van twee verschillende bronnen aan elkaar gekoppeld. Het gaat om gegevens met betrekking tot arbeidsmarktsucces van *afgestudeerde werkzoekenden* per opleiding en gegevens over het *aantal studenten over alle jaren* per opleiding. Het percentage studenten dat werk vindt, kan per jaar sterk verschillen afhankelijk van de conjunctuur of tijdelijke omstandigheden binnen het ROC (denk bijvoorbeeld aan slechte leraren). Daarom is het niet helemaal zuiver om het percentage van een afstudeerjaar te gebruiken als percentage voor de gehele opleiding (dus voor alle jaren).

Daarnaast speelt er een probleem met de doorstroom van studenten. Omdat een groot deel van de studenten na een mbo-opleiding verder studeert, meet de eerste dataset slechts een deel van de populatie die je wil onderzoeken. Daardoor kan er een *bias* optreden, omdat het voor de hand ligt dat het de betere studenten zijn die doorstromen (en dus niet als afgestudeerde werkzoekende geregistreerd worden). Daardoor ligt het werkelijke werkloosheidspercentage voor veel opleidingen waarschijnlijk lager dan het percentage dat hier gebruikt is.

Een manier om dit te ondervangen, is door naar data van individuele studenten te kijken, in plaats van naar data per opleiding. Deze data zijn beschikbaar (er worden ieder jaar op grote schaal enquêtes afgenomen onder studenten) maar moeilijk te koppelen aan instellingen.

Methodologie

In dit onderzoek is gebruik gemaakt van OLS-regressie. In het hoofdstuk 'methodologie' is al kort bij de beperkingen hiervan stilgestaan. Omdat de afhankelijke variabelen een percentage is, is het denkbaar dat het regressiemodel waarden voorspelt die niet mogelijk zijn in de praktijk.

Een alternatief is het voorspellen van de kans op werk per student aan de hand van de gegevens van de opleiding. Dat kan als de bovenstaande gegevens beschikbaar zijn. Met een logistische regressie kan dan de kans op werk voor een specifieke student voorspeld worden, gegeven de grootte van de opleiding en andere relevante factoren.

Conclusies

Deze scriptie begon met de vraag wat de invloed is van opleidingsgrootte op de kans op werk op mbo4. Er werden twee mogelijke manieren besproken waarop grootte van de opleiding kan doorwerken: via schaalvoordelen en via het *class size-effect*. Daarna is door middel van het mergen van twee datasets en verschillende OLS-regressies getest of grootte van de opleiding van invloed is op de kans op werk. Uit vier van de zes tests is gebleken dat die invloed niet significant is. Als het wel significant was, verdween die significantie als controlevariabelen werden toegevoegd. De hoofdvraag van de scriptie luidde:

“Wat is de invloed van de grootte van een mbo4-opleiding op de kans op een baan voor een student?”

Hoewel deze scriptie een aantal beperkingen kent die uitgebreid aan bod zijn gekomen, kan toch met enige zekerheid geconcludeerd worden dat de grootte van de opleiding in ieder geval niet van doorslaggevend belang is voor de kans op een baan voor studenten aan het mbo op niveau 4.

Er zijn suggesties gedaan voor verder onderzoek, met alternatieve methoden of met betere data. Daarnaast is het interessant om te kijken naar factoren die bijdragen aan kans op een baan maar die niet aan bod zijn gekomen in deze scriptie. In alle regressiemodellen van deze scriptie was de verklarende kracht zeer laag, er zijn dus nog voldoende mogelijkheden. Er zou bijvoorbeeld gekeken kunnen worden naar beschikbaarheid van stages, kwaliteit van docenten of specifiek naar welke sectoren succesvol zijn.

Referenties

Literatuur

Blank, J. L. T., van Heezik, A. A. S. (2015). *Productiviteit van het overheidsbeleid. Deel 1: Het Nederlands onderwijs 1980 – 2012*. Den Haag/Delft: Centrum voor Innovaties en Publieke Sector Efficiëntie Studies.

Chingos, M. M. (2013). Class size and student outcomes: research and policy implications. *Journal of Policy Analysis and Management*, 2, 411-438.

Gelderblom, A., Gravesteijn, J., Stegehuis, B. J., Vleeschouwer, E. (2016). *MBO-instellingen en de arbeidsmarktrelevantie van het opleidingsaanbod op niveau 2*. SEOR, Rotterdam.

Harfitt, G. J., Tsui, A. B. M. (2015). An examination of class size reduction on teaching and learning processes: a theoretical perspective. *British Educational Research Journal*.

Koning, J. de, Gelderblom, A. en Gravesteijn, J. (2010) Techniek: exact goed? Het keuzeprocess van allochtone en autochtone leerlingen in het (V)MBO verklaard. SEOR, Rotterdam.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*.

Meng, C. Sijbers, E. (2017). *Arbeidsmarktrelevantie mbo niveau 2*. ROA, Maastricht.

Data

Dienst Uitvoering Onderwijs (DUO). (2018, 4 mei). *Aantal deelnemers in het mbo*. Gedownload van https://duo.nl/open_onderwijsdata/databestanden/mbo/onderwijsdeelnemers/

Centraal Bureau voor Statistiek (CBS). (2017, 6 augustus). *Arbeidsmarktkenmerken mbo-schoolverlaters 2014/'15*. Gedownload van <https://www.cbs.nl/nl-nl/maatwerk/2017/33/arbeidsmarktkenmerken-mbo-schoolverlaters-2014-15>

Appendix

Methodologie

Tabel 1: Koenker en Breusch - Pagan test (H_0 : homoscedasticiteit)

Tests voor Heteroscedasticiteit		
	Test score	Significantie
Koenker	3,123	0,772
Breusch - Pagan	4,683	0,305

Tabel 2: Tolerance- en VI-factor per variabele

Multicolineariteit		
	Tolerance Factor	VI-Factor
Aantal studenten	0,927	1,079
Type bbl	0,937	1,067
Sector economie	0,632	1,583
Sector groen	0,850	1,177
Sector techniek	0,635	1,576
Randstad	0,992	1,008

Tabel 3: Durbin-Watson test

Autocorrelatie: Durbin-Watson score per vergelijking						
	(1)	(2)	(3)	(4)	(5)	(6)
Score	0,270	0,275	0,270	0,275	0,272	0,275