

September 2016

STW Evaluatie:

Bijlage Econometrische Effectmeting

STW Evaluatie:

Bijlage Econometrische Effectmeting

technopolis _{group} September 2016

Matthias Ploeg MSc

Dr. Frank Zuijdam

Met dank aan de klankbordgroep econometrische effectmeting:

Prof. Dr. Carl Koopmans, voorzitter (VU/SEO)

Dr. Piet Donselaar (Ministerie van Economische Zaken)

Ir. Fred Couzy (Ministerie van Economische Zaken)

Dr. Marja Dijksterhuis (STW)

Inhoudsopgave

1	Introductie	3
1.1	Context en doelstelling	3
1.2	Uitdaging Effectmeting: selectiebias	3
1.3	Overzicht methodologische annex.....	3
2	Methodologie: Maatstaven en Schattingsmethodes	4
2.1	Keuze maatstaven wetenschappelijke kwaliteit en utiliteit	4
2.2	Keuze schattingsmethoden	5
2.3	Afbakening en beperkingen	7
3	Specificatie van Schattingsmethoden	8
3.1	Beschrijving van de onderzoekspopulatie	8
3.2	Beschrijvende analyse van de maatstaven	9
3.3	Verschil-in-verschil analyse.....	10
3.4	Specificatie regressie-discontinuïteitsanalyse.....	12
3.4.1	Juryscores, cross-overs en no-shows	12
3.4.2	Trendbreuk rondom afkappunt.....	13
3.4.3	Analyse polynomen	14
3.4.4	Keuze bandbreedte	14
3.4.5	Overzicht modelspecificaties	15
3.5	Specificatie fixed-effects / random-effects (panel) analyse	15
3.5.1	Fixed-effects vs Random effects	15
3.5.2	Analyse succesfactoren & Gemeenschappelijke trendanalyse.....	16
3.5.3	Overzicht specifieke specificatie.....	17
4	Resultaten	18
4.1	Overzicht resultaten.....	18
4.2	Aanvullende analyse effectperiode	18
5	Discussie en Conclusie	23

1 Introductie

1.1 Context en doelstelling

Deze econometrische annex geeft een overzicht van de resultaten van de econometrische analyse die is uitgevoerd in het kader van deze STW-evaluatie. De econometrische effectmeting is deels een (aangepaste) herhaling alsmede een uitbreiding van de studie die in 2013 door het CPB is gepubliceerd.¹ De econometrische effectmeting heeft als doel in kaart te brengen wat de *causale* effecten zijn van het verkrijgen van een STW-beurs (in dit geval een subsidie via het Open-Technologie Programma (OTP) of een Thematisch Programma (TP) op zowel wetenschappelijke als utiliteitsmaatstaven.

Een uitgebreide samenvatting van de context van deze evaluatie, alsmede een beschrijving van de het Open-Technologie Programma en de Thematische Programma's is beschikbaar in het hoofdrapport van deze evaluatie.

1.2 Uitdaging Effectmeting: selectiebias

De voornaamste uitdaging bij het vaststellen van een dergelijk effect is een zogenaamde 'selectiebias', dat wil zeggen dat succesvolle aanvragers fundamenteel anders zouden kunnen zijn dan niet-succesvolle aanvragers, waardoor een stijgende productiviteit wordt aangezien voor een effect, terwijl deze stijging mogelijk ook (deels) had plaatsgevonden zonder de beurs. Aangezien er vanzelfsprekend geen experimentele technieken (willekeurige toekenning) zijn gebruikt bij de uitvoering van de STW-Beurzen, zoals gebruikelijk in exact en medisch, wetenschappelijk onderzoek, zal hier gebruik moeten worden gemaakt van *quasi-experimentele methoden* om een effect te meten. Het onderzoeksteam heeft hierbij de richtlijnen voor econometrische effectmeting van de Commissie Theeuwes² opgevolgd. Deze zelfstandige annex econometrische effectmeting STW geeft een gedetailleerd overzicht van de gebruikte methodes.

1.3 Overzicht methodologische annex

Deze methodologische annex is als volgt ingedeeld:

- **Sectie 2** bespreekt de keuze voor maatstaven en schattingsmethodes.
- **Sectie 3** bespreekt in detail de beschrijvende gegevens van de maatstaven, en bespreekt in detail de specifieke specificaties van de schattingsmethoden
- **Sectie 4** geeft de resultaten weer.
- **Sectie 5** bevat de discussie en conclusie van deze resultaten.

¹ CPB Discussion Paper 248, 'Effects of Research Grants on Scientific Productivity and Utilisation' (2013). Debby Lanser & Ryanne van Dalen.

² 'Durf te meten', Eindrapport Expertwerkgroep Effectmeting (november 2012).

2 Methodologie: Maatstaven en Schattingsmethodes

2.1 Keuze maatstaven wetenschappelijke kwaliteit en utiliteit

Aangezien de doelstelling van STW en de STW-beurzen zowel het bevorderen van Wetenschappelijke Kwaliteit als Utiliteit (toepassing van kennis) is, is het voor een relevante effectmeting belangrijk om ook relevante maatstaven in beide domeinen te kiezen. Voortbouwend op de CPB-studie is een redelijk breed palet aan maatstaven voor wetenschappelijke kwaliteit (WK) en Utiliteit (UT) gekozen. Een belangrijke beperking hierbij is dat enkel data op onderzoeker (aanvrager) niveau beschikbaar was, niet op projectniveau of op bedrijfsniveau (van betrokken bedrijven). Dit betekent dat de gehele analyse over het algemeen door een zeer ‘wetenschappelijke lens’ wordt bekeken (zie ook discussie van beperkingen in sectie 2.3). De resultaten vallen dan ook beter te interpreteren als een effectmeting van door STW-gesteunde *onderzoekers (aanvragers)* dan een effectmeting van door STW gesteunde *onderzoeksprojecten*.

In navolging van de discussie in het CPB-onderzoek alsmede gesprekken met STW-staf is gekozen – met dank aan een inmiddels een vrij rijke en ‘lange’ dataset’ - niet alleen te focussen op de effecten tijdens de implementatie van het project, maar ook de effecten in de periode daarna (de ‘resultaatperiode’). Dit soort type onderzoeksprojecten, gesteund door promovendi, hebben namelijk een vrij lange aanloopperiode plus een vertraging tussen effect (onderzoek afgerond) en het meten van het effect (onderzoekspaper gepubliceerd). Voor elke maatstaf zijn dus twee varianten meegenomen: één met de geaggregeerde effecten in de implementatieperiode (jaar 1 tot en met jaar 4 na toekenning) en één met de geaggregeerde effecten in de ‘resultaatperiode’ (jaar 5 tot en met jaar 8 jaar na toekenning).

Tabel 1 Overzicht maatstaven

WK/UT	Definitie	Bron (*= nieuw t.o.v. CPB studie)
Publicaties (WK)	Aantal publicaties & reviews in Web of Science (1992-2015).	Web of Science (CWTS Analyse)
Citatiescore (WK)	Aantal ontvangen citaties genormeerd op 250 Journal categorieën in WOS (1992-2015). Citaties beperkt tot 4 jaar na publicatie.	Web of Science (CWTS Analyse)
	Aantal ontvangen citaties genormeerd op ongeveer 4000 clusters van publicaties (Leiden Ranking). Citaties beperkt tot 4 jaar na publicatie	Web of Science (CWTS Analyse)
Publiek-Private Publicaties (UT/WK)	Aantal publicaties met een bedrijf als een van de coauteurs in de adresgegevens.	Web of Science (CWTS Analyse)
Aandeel Publiek-Private Publicaties (UT/WK)	Aandeel van publiek-private publicaties in het geheel van de publicaties	Web of Science (CWTS Analyse)*
Octrooiaanvragen (UT)	Aantal octrooiaanvragen bij het EPO-patentbureau (-2015)	PATSTAT Online
Octrooicitaties (UT)	Aantal citaties van de octrooien aangevraagd door de STW-aanvragers.	PATSTAT Online*
NWO Beurzen: Talent Programma's (WK)	Positief (1) indien aanvrager 1 of meer succesvolle aanvragen heeft gedaan als hoofdaanvrager bij de volgende NWO/STW programma's: Vernieuwingsimpuls, Rubicon, Aspasia, Mozaïek	NWO Database*

**NWO Beurzen:
Excellentie
Programma's
(WK)**

Positief (1) indien aanvrager 1 of meer succesvolle aanvragen als hoofdaanvrager heeft gedaan bij de volgende NWO/STW programma's: Vrije competitie, Open Competitie, TOP, Echo, Open Programma

NWO Database*

2.2 Keuze schattingsmethoden

In navolging van de aanbevelingen gedaan door de Commissie Theeuwes, zijn de volgende vijf methoden overwogen om het probleem van de selectiebias te adresseren. De algemene voor- en nadelen van deze methoden worden uitvoerig besproken in het samenvattend verslag van deze Commissie en zullen hier niet worden herhaald³.

- Verschil-in-verschil analyse.
- Regressie discontinuïteit-analyses (RDD)
- Propensity-Score Matching (PSM)
- Instrumentele variabelen (IV)
- Fixed-Effects Analyse (FE)

De methode instrumentele variabelen viel helaas al snel af aangezien er in deze dataset geen goed instrument voorhanden was. Ook Propensity Score-Matching bleek lastig implementeerbaar aangezien de beschikbare achtergrondvariabelen slechts een zeer klein beperkt voorspellend effect hebben op uiteindelijke selectie. Gelukkig beschikt deze dataset zowel over jureringsgegevens die een RDD-methode mogelijk maken alsmede een 'panel'datastructuur die een fixed-effects analyse mogelijk maken. De FE/RE methodes zijn in alle opzichten superieur ten opzichte van verschil-in-verschil analyses, omdat deze een vergelijkbare aanpak kennen maar meer informatie kunnen meenemen. Vanwege de intuïtieve interpretatie wordt de verschil-in-verschil methode echter wel reeds besproken in sectie 2.3. Dit bracht de keuze voor de schattingsmethodes terug tussen de RDD en de FE-methodes, waarbij in principe de RDD-methode de voorkeur heeft omdat deze aanpak het meest dicht bij een echte experimentele methode komt en daarmee minder assumpties nodig heeft voor een nauwkeurige schatting van de effecten. De voorlopende CPB-studie maakte ook gebruik van deze methode.

Er zijn echter een aantal beperkingen aan het gebruik van de RDD-methode in deze specifieke context. De eerste beperking is dat juryscores slechts voor een beperkt deel van de Thematische Programma's beschikbaar zijn (zie onderstaande tabel), omdat niet voor alle Thematische Programma's voor de zelfde jurysystematiek is gekozen.

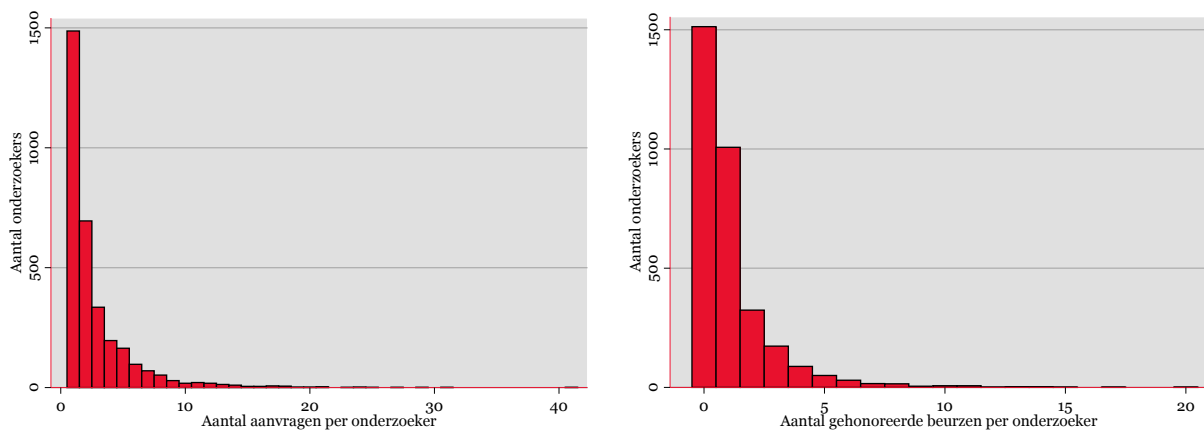
Tabel 2 Overzicht beschikbaarheid Jurydata

	<i>OTP</i>	<i>TP</i>	<i>Totaal</i>
Geen Jurydata	195	320	515
Wel Jurydata	2313	375	2688
Totaal	2508	695	3203

³ Durf te meten', Eindrapport Expertwerkgroep Effectmeting (november 2012).

Een andere uitdaging is de structuur van de onderzoekspopulatie waarbij onderzoekers meerdere aanvragen en meerdere toekenningen kunnen krijgen (zie figuur beneden). Het komt vrij regelmatig voor dat onderzoekers niet alleen meerdere aanvragen indienen, maar ook meerdere beurzen krijgen toegekend, soms ook binnen 1 jaar. Één onderzoeker diende 42 aanvragen in waarvan er 21 werden goedgekeurd binnen de periode 1996-2011. Hierdoor komen dezelfde onderzoekers meerdere malen voor in de data op *onderzoeker-aanvraag* niveau.

Figuur 1 Overzicht aanvragen en honoreringen per onderzoeker



Belangrijk hierbij is dat de maatstaven echter alle gemeten zijn op *onderzoekersniveau*, terwijl de data rondom aanvragen beschikbaar is op *onderzoeker-aanvraag niveau*. Een nadeel van het gebruik van individuele aanvragen terwijl de data op onderzoekersniveau is gemeten, is dat er niet is voldaan aan de Gauss-Markov voorwaarde van het ontbreken van correlatie tussen de fouttermen van de observaties (omdat dezelfde data meerdere malen als unieke observatie wordt meegenomen), waardoor OLS niet meer een ‘beste lineaire *unbiased* schatter (‘BLUE’) is⁴. Een ander probleem met het gebruik van de data op *onderzoeker-aanvraag* niveau is dat er een risico is dat effectperiodes overlappen wanneer een aanvrager meerdere beurzen binnen een bepaalde periode ontvangt. Hierdoor vervaagt een duidelijk conceptueel onderscheid tussen ‘voor’ interventie en ‘na’ interventie. Er ontstaat hierdoor een risico van ‘dubbel tellen’ wanneer meerdere beurzen in dezelfde periode worden ontvangen, waarbij elke beurs het geaggregeerde effect ‘voor eigen rekening’ neemt⁵, en daarmee wordt de effectmeting onbetrouwbaar. Gebruikmakend van de paneldataset is het eenvoudig te onderzoeken in welke mate het geconstateerde probleem van effectoverlap zich in de praktijk manifesteert. Uit een analyse van de data blijkt dat wanneer een periode van tot 8 jaar na toekenning (dus hiermee de twee type effectperiodes omvattend) wordt aangehouden, er 12% van de onderzoekers onder 2 of meer effectperiodes tegelijk valt. In totaal 2% van de onderzoekers valt zelfs onder 4 periodes tegelijk, met als maximum een drietal onderzoekers die op een moment onder de effectperiode van 8 beurzen tegelijk vallen. Dit laat zien dat deze overlap inderdaad voor een substantiële groep het geval is.

⁴ Voor een diepere beschrijving van de Gauss-Markov assumpties zie ook Greene, W. (2011). *Econometric Analysis*, 7th edition, Pearson, New York

⁵ Sandler, R. & Sandler, M (2013). *Multiple Event Studies in Public Finance and Labor Economics: A Simulation Study with Applications*.

Deze beperkingen maken dat we in deze econometrische effectmeting kiezen voor een analyse op *onderzoekersniveau*. Dit betekent dat er voor een RDD-OLS regressie slechts kan worden gewerkt met 1 aanvraag per onderzoeker, omdat er concrete effectperiodes met elkaar moeten worden vergeleken. Hierdoor vallen alle onderzoekers met meerdere honoreringen per definitie af, wat dan weer een groot nadeel is van deze aanpak.

Omdat bij een dergelijke dataset met ‘meerdere interventies’ over het algemeen wordt geadviseerd om een fixed-effects aanpak te gebruiken⁶, en omdat de RDD-analyse dus een substantieel deel van de data niet kan meenemen, is gekozen om beide schattingsmethoden naast elkaar uit te voeren. Hierbij is de RDD-aanpak in principe conceptueel te prefereren, terwijl de fixed-effects analyse het mogelijk maakt het effect over de hele onderzoekspopulatie te onderzoeken én het effect van meerdere toekenningen te kunnen onderzoeken.

2.3 Afbakening en beperkingen

Het is zeer belangrijk de afbakening en beperkingen van deze studie in acht te nemen bij het interpreteren van de resultaten:

- Deze analyse **betreft enkel OTP- en TP-beurzen die vielen in de periode 1996-2011**. In totaal betrof dit 8800 unieke aanvragen van ongeveer 3600 unieke onderzoekers (aanvragen zijn meestal gezamenlijk met 2 à 3 onderzoekers).
- De opgevoerde maatstaven voor **wetenschappelijke kwaliteit, collegiaal getoetste publicaties en gewogen citatiescores zijn over het algemeen zeer valide** en worden algemeen gebruikt om kwaliteit en hoeveelheid van wetenschappelijke productiviteit te meten. Er dient hierbij te worden opgemerkt dat in bepaalde subdisciplines (bijvoorbeeld toegepaste bouwkunde) andere resultaten (zoals modellen) belangrijker zijn dan publicaties.
- **De opgevoerde maatstaven voor utiliteit geven slechts een beperkt beeld** van de totale utiliteitsimpact van het STW-onderzoek. Octrooien en publiek-private publicaties zijn slechts voor een beperkt deel van de disciplines relevant en alle verwachte impacts op betrokken bedrijven (of maatschappelijke organisaties) blijven volledig buiten beeld vanwege een exclusieve focus op onderzoekers en geen toegang tot bedrijfsgegevens.
- **De opgevoerde maatstaven zijn erg gericht op het meten van een ‘gemiddeld’ effect op kwantiteit**. Omdat er geen data op projectniveau beschikbaar is, kunnen enkel data op onderzoekersniveau worden vergeleken. De schattingsmethoden die hiervoor algemeen worden gebruikt zijn geschikt om een gemiddeld effect te kunnen berekenen. Mocht er sprake zijn van een klein aantal zeer excellente onderzoeksprojecten met een zeer grote impact, dan komt een dergelijk gedifferentieerd effect niet goed terug in deze analyse.
- De beschreven effecten moeten worden **geanalyseerd als ‘additionele’ effecten bovenop het bestaande andere onderzoeksinstrumentarium**. Hiermee wordt het effect dus deels onderschat, omdat afgewezen onderzoekers elders financiering kunnen hebben verkregen. Wanneer de gehele STW-financiering zou wegvallen, zou dit zeer waarschijnlijk niet kunnen worden opgevangen elders in het systeem. Gezien de vrij unieke functie van deze STW-beurzen in het Nederlandse systeem en de lage overlap met andere NWO-beurstoekenningen verwachten wij echter dat deze verstoring niet overmatig groot is.

⁶ Ibid.

3 Specificatie van Schattingsmethoden

3.1 Beschrijving van de onderzoekspopulatie

Teneinde een goede specificaties van de schattingsmethodes te kunnen selecteren, is het van belang om een goed inzicht te hebben in de onderzoekspopulatie (deze sectie) en de specificatie van de gekozen maatstaven (volgende sectie).

Het uitgangspunt was een bestand met 8992 unieke onderzoeker-aanvraag combinaties in de periode 1996 tot en met 2011. Met onderzoeker-aanvraag combinatie bedoelen we één specifieke onderzoeker binnen één specifieke aanvraag; een onderzoeker kan immers meerdere aanvragen doen en een aanvraag kan worden gedaan door meerdere onderzoekers. Deze periode is gekozen omdat er voor 1996 slechts een zeer beperkt aantal beurzen per jaar werden uitgereikt, en het voor onderzoeksprojecten in de periode na 2011 nog niet mogelijk is de volledige effecten in kaart te brengen. Dit bestand van 8992 onderzoeker-aanvraag combinaties is uiteindelijk teruggebracht naar 3608 unieke onderzoekers, waarvan er voor 3202 data beschikbaar was (onderzoekers waarvan geen enkele publicatie is gevonden zijn uitgesloten van de analyse). Onderstaande tabel geeft de belangrijkste achtergrondvariabelen van deze onderzoekers weer.

Tabel 3 Overzicht maatstaven

Indicator	N	gemiddelde	min	max
% Met een (of meer) toekenning	3202	53%		
Aantal gehonoreerde OTP aanvragen	3202	0,63	0	15
Aantal gehonoreerde TP aanvragen	3202	0,46	0	10
Aantal OTP aanvragen	3202	1,86	0	33
Aantal TP aanvragen	3202	0,90	0	16
Aantal aanvragen als Mede-Aanvrager	3202	1,69	0	24
Aantal aanvragen als Hoofd-Aanvrager	3202	1,08	0	20
Totaal aantal honoreringen	3202	1,09	0	20
Totaal aantal aanvragen	3202	2,77	1	41
% Met slechts 1 aanvraag	3202	46%		
% Vrouwen	3035	9%		
% Met heraanmelding zelfde aanvraag	3202	2%		
% met meer dan 1 toekenning	3202	23%		

Zoals besproken in sectie 2, is er dus gekozen om het onderzoek uit te voeren op onderzoeker-niveau en niet op onderzoeker-aanvraag niveau. Bij onderzoekers die dezelfde aanvraag meerdere malen indienden (herindieningen), is de laatste aanvraag (met bijbehorende juryscore en resultaat) meegenomen. Wel is er een controle-indicator gecreëerd voor deze gevallen. Voor onderzoekers die meerdere niet-succesvolle aanvragen hebben ingediend is de *laatste* aanvraag meegenomen als referentiepunt (enkel relevant voor de RDD-analyse).

3.2 Beschrijvende analyse van de maatstaven

Een beschrijvende analyse van de maatstaven is weergegeven in het onderstaande tabel (op basis van de complete paneldata). Over het algemeen is alle data beschikbaar tussen 1992-2015, waardoor voor het grootste deel van de populatie een volledige set geaggregeerde periodes kan worden samengesteld ('voor', 'implementatie' en 'resultaat'⁷). Daarnaast is het belangrijk om te vermelden dat voor de maatstaven octrooien en NWO-beurzen ervan uit is gegaan dat geen gevonden data betekent dat er geen productiviteit was⁸. Voor publicaties is er echter vanuit gegaan dat wanneer er totaal geen publicaties gevonden zijn dit komt omdat deze onderzoekers niet juist geïdentificeerd zijn, niet omdat er geen productiviteit was. Deze onderzoekers zijn uitgesloten van de populatie. Daarnaast is gecontroleerd voor de 'start' van carrières door de jaren dat onderzoekers nog geheel niet actief waren (gemeten aan publicaties) niet mee te nemen in de analyse (hiermee daalt het beschikbare datapunten van 79.080 tot 69.053)

Tabel 4 Beschrijvende statistieken maatstaven; per JAAR

Maatstaf	Observaties (uit 79.080)	Gemiddelde	Std. Dev.	Min	Max
Publicaties	69.053	4,31	5,56	0	133
Publiek-Private Publicaties	69.053	0,41	1,01	0	25
Aandeel Publiek-Private Publicaties	69.053	0,10	0,23	0	100
Citatiescore (1)	55.216	1,45	2,17	0	240
Citatiescore (Leiden Ranking)	39.615	1,45	1,61	0	47
EPO Octrooiaanvragen	69.053	0,10	0,73	0	44
EPO Octrooicitaties	69.053	0,22	4,16	0	528
NWO-Talentbeurzen	69.053	0,40%	N.v.t.	0	1
NWO-Open Competitie beurzen	69.053	0,85%	N.v.t.	0	1

⁷ Hierbij is gekozen om voor deze periode's eerst het gemiddelde uit te rekenen voor elke maatstaf, waarna deze met vier is vermenigvuldigd. Hierdoor kunnen ook indicatief voor effectperiodes worden geschat wanneer er geen gegevens zijn voor de volledige vier jaar.

⁸ Aangezien het waarschijnlijk is dat de datakoppelingsprocedure niet geheel foutloos is verlopen (bijvoorbeeld door naamwijzigingen, of verkeerd gespelde namen), is het dus waarschijnlijk dat het aantal octrooien en octrooicitaties, alsmede het aantal NWO-beurzen licht is onderschat. Er is echter geen reden om aan te nemen dat hier een structureel verschil in zou zitten tussen succesvolle en niet-succesvolle aanvragers.

Uit analyse blijkt dat de correlaties tussen de maatstaven vrij beperkt zijn (wat een voorwaarde is voor een goede schatting om een eventueel probleem met multicollineariteit te voorkomen), behalve voor publiek-private publicaties en publicaties (correlatiecoëfficiënt van 0,47). Daarom is de toevoeging van een ‘aandeel’ publiek-private publicaties een goede aanvulling.

3.3 Verschil-in-verschil analyse

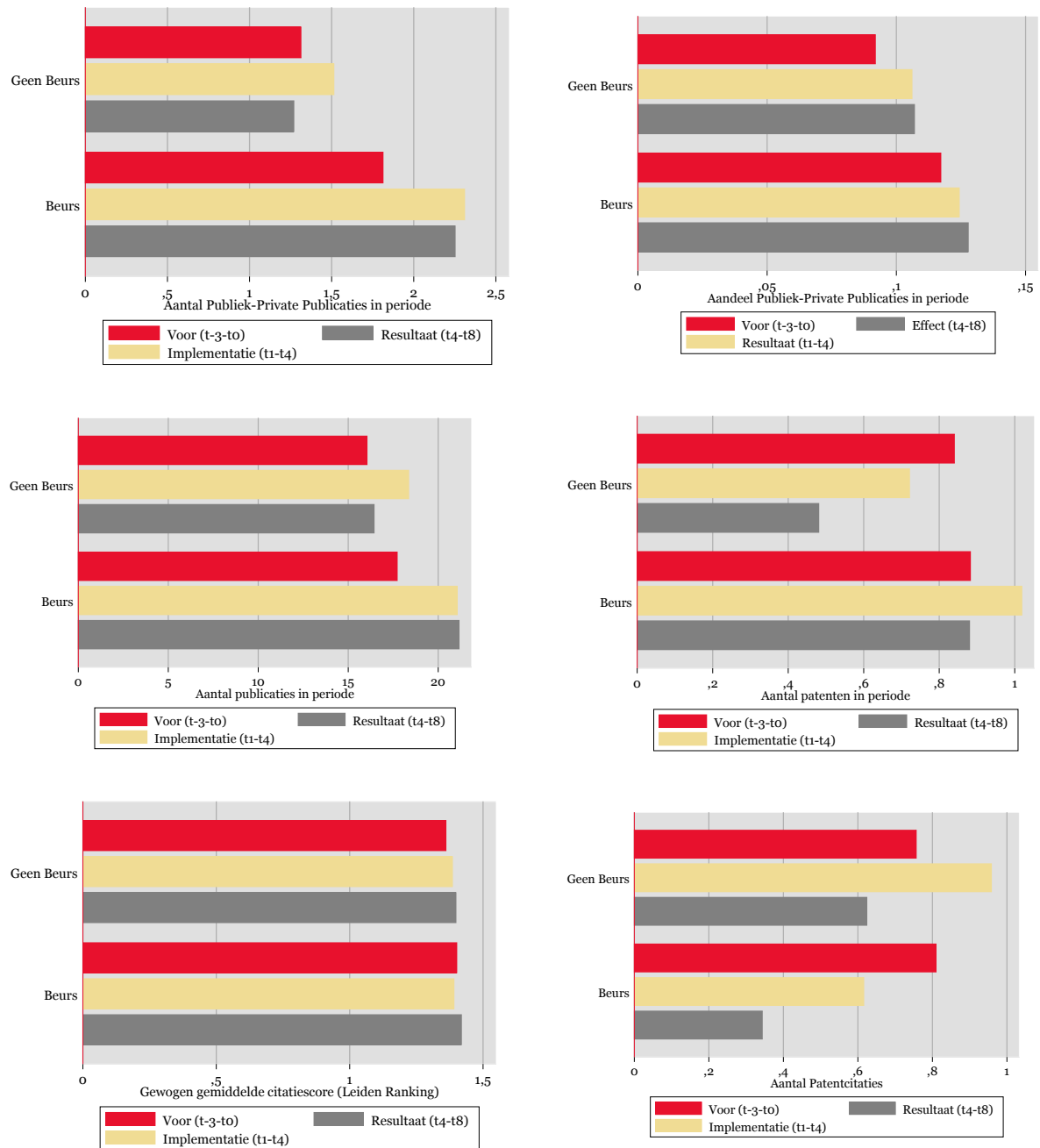
Om een goed idee te krijgen van de ordes van grootte van de verschillende effecten is een verschil-in-verschil analyse uitgevoerd. In de praktijk betekent dit dat we het verschil van bijvoorbeeld het aantal publicaties in de periode voor (t-3 tot en met to) aftrekken van het aantal publicaties in de periode na toekenning voor zowel deelnemers als niet-deelnemers, en daarna het verschil tussen de deelnemersgroep en de niet-deelnemers berekenen (zie kolom ‘effect’ in onderstaande tabel). Ook is via een eenvoudige t-toets gekeken of dat laatste verschil tussen deelnemers en niet-deelnemers significant was. Deze verschil-in-verschil analyse is uitgevoerd met hetzelfde databestand dat voor de RDD-analyse gebruikt wordt, omdat daarbij duidelijke ‘Voor’ en ‘Na’ periodes te onderscheiden zijn. De resultaten van deze analyse zijn gepresenteerd in onderstaande tabel en figuur.

Wat opvalt is dat er een significant verschil-in-verschil is voor publicaties en publiek-private publicaties, alsmede octrooi. Het aandeel publiek-private publicaties verandert echter niet significant. Ook is er een klein negatief verband geconstateerd tussen toekenning van een STW-beurs en het succesvol aanvragen van een NWO-Talentbeurs. Een Fixed Effect/Random Effects analyse heeft methodologisch gezien de voorkeur (omdat hierbij meerdere beurzen per onderzoeker kunnen worden meegenomen, zie ook discussie in Sectie 2.2), maar een verschil-in-verschil analyse geeft wel direct een intuïtief eerste inzicht in de orde van grootte van de verschillen tussen de groepen.

Tabel 5 Verschil-in-verschil analyse (op basis van de RDD data)

	Voor (t-3 t/m to)		Implementatie (t1 t/m t4)		Resultaat (t5 t/m t8)		Implementatie (verschil-in-verschil)		Resultaat (verschil in verschil)	
	Nee	Ja	Nee	Ja	Nee	Ja	‘Effect’	Sig	‘Effect’	Sig
Publicaties	16,05	17,72	18,36	21,06	16,44	21,16	1,02	**	3,05	***
Publiek private publicaties	1,31	1,81	1,51	2,31	1,27	2,25	0,30	***	0,48	***
Aandeel publiek-private publicaties	0,09	0,12	0,11	0,12	0,11	0,13	0,01		0,01	
EPO Octrooi Citaties	0,76	0,81	0,96	0,62	0,62	0,34	-0,40		-0,33	
EPO Octrooi Aanvragen	0,84	0,88	0,72	1,02	0,48	0,88	0,25	**	0,36	
Gemiddelde citatie score	1,34	1,44	1,32	1,41	1,34	1,36	-0,01		-0,01	
Gemiddelde citatie score (2)	1,36	1,40	1,39	1,39	1,40	1,42	-0,01		-0,04	
% met een NWO Talent Beurs toekenning	2%	3%	1%	2%	1%	1%	-1%	*	-1%	**
% met een NWO Open Competitie toekenning	3%	4%	4%	4%	2%	3%	-1%		1%	

Figuur 2 Verschil-in-Verschil analyse (onder de populatie met maximaal één succesvolle aanvraag)



3.4 Specificatie regressie-discontinuïteitsanalyse

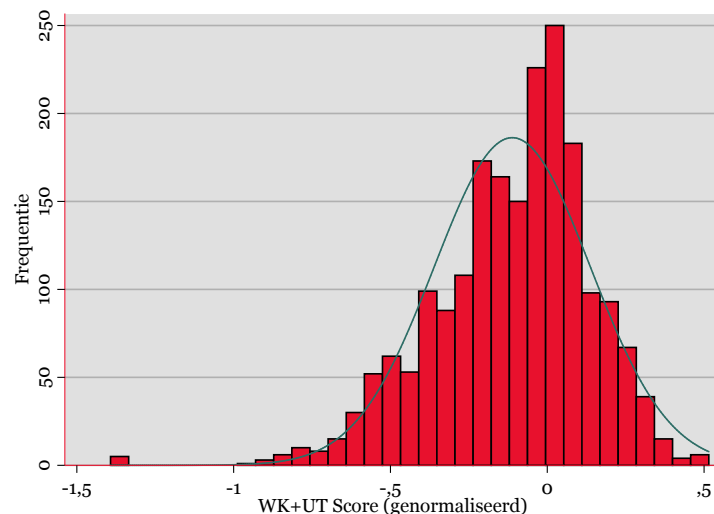
Om de regressie-discontinuïteit analyse (RDD) goed uit te voeren is het belangrijk om de karakteristieken van de jurysystematiek en de interactie met de juryscores goed in kaart te brengen om eventuele versturende factoren te kunnen identificeren en waar mogelijk te adresseren. Hiervoor wordt een vergelijkbaar stappenplan gevolgd als bij de Eureka-Eurostars Evaluatie 2011-2015⁹.

3.4.1 Juryscores, cross-overs en no-shows

Bij het Open Technologie Programma is reeds sinds lange tijd vrij consistent een jurysystematiek gebruikt om de beperkte middelen te verdelen onder de beste selectie voorstellen. Voorstellen worden beoordeeld door juryleden die onafhankelijk van elkaar elk voorstel binnen een specifieke ronde twee scores geven: een score voor wetenschappelijke kwaliteit (WK) en een score voor utiliteit (UT), oftewel toepasbaarheid. Scores lopen van 1 (uitmuntend) tot 9 (zeer slecht). De scores worden door STW opgeteld (beide scores wegen even zwaar) tot een integrale juryscore. Op basis van deze som wordt een rangorde opgesteld. Deze rangorde wordt voorgelegd aan het STW-bestuur voor een definitief besluit¹⁰.

Deze jurysystematiek leent zich in principe uitstekend voor de toepassing van de RDD-systematiek, met name omdat deze zo consistent is toegepast en het opstellen van de rangorde inderdaad zorgt voor een enigszins 'toevallige' toekenning rondom het afkappunt. De figuur hieronder toont een histogram waarin de genormaliseerde (om absolute verschillen tussen jaren heen te schalen en het afkappunt altijd op 0 te hebben liggen) juryscore is weergegeven. Om de interpretatie intuïtiever te maken is de juryscore hierbij ook omgedraaid, waarbij hoger beter is.

Figuur 3 Genormaliseerde Juryscore



Juryscores zijn niet geheel deterministisch: er zijn een substantieel aantal 'crossovers' en een klein aantal 'no-shows'; zie de tabel hieronder. Een cross-over is een voorstel dat een score had onder het afkappunt maar uiteindelijk toch is toegekend. Een 'no-show' is daarentegen juist een voorstel dat wel een score boven het afkappunt had maar uiteindelijk geen beurs heeft gekregen. Een dergelijk fenomeen betekent dat een zogenaamde 'fuzzy' regressie discontinuïteitsanalyse moet worden toegepast (zie verder beneden). In principe heeft een 'fuzzy'-aanpak geen consequenties voor de nauwkeurigheid van de effectmeting, maar hierbij is wel de voorwaarde dat het voorkomen van een

⁹ Prince, Y. et. al (2014). Evaluatie Eureka en Eurostars 2012-2015. Panteia

¹⁰ Voor een gedetailleerd overzicht van de jurysystematiek, zie: STW (2013) Richtlijnen voor Juryleden. Open Technologie Programma

cross-over of een no-show ‘willekeurig’ is, dat wil zeggen niet gerelateerd is aan waarneembare of niet-waarneembare karakteristieken van de aanvrager.

Tabel 6 *Overzicht cross-overs en no-shows*

	<i>Geen toekenning</i>	<i>Wel toekenning</i>
Genormaliseerde score < 0	1174	96
Genormaliseerde score > 0	25	809

Naar aanleiding van de constatering van de aanwezigheid van cross-overs en no-shows heeft STW onderzocht waar deze resultaten door te verklaren zijn. De voornaamste reden voor het vinden van een cross-over (of no-show) is dat voor 2008 voor het OTP-programma en voor een kleine selectie van Thematische Programma’s het STW-bestuur het uiteindelijke oordeel velde over de ranglijst die op basis van juryscores was samengesteld. Voor de Thematische Programma’s werd het bestuur hierbij geadviseerd door een inhoudelijke programmacommissie, voor OTP werd deze beslissing direct door het Bestuur zelf genomen. Hier is in bepaalde gevallen – gemotiveerd – gekozen om tot een andere beslissing te komen. De voornaamste reden hierbij was de coherentie binnen het programma en de verdeling over de subthema’s van STW. In het algemeen is het STW Bestuur gerechtigd om de keuze van projecten aan te passen op basis van beleidsafwegingen. Belangrijk hierbij is te melden dat deze afwegingen niet projectinhoudelijk mogen zijn, d.w.z. de subjectieve kwaliteit van een projectvoorstel en achterliggende onderzoeker niet mee mogen worden genomen in deze beslissing. Dit laatste is met name van belang om de assumptie dat onderzoekers net boven en onder het afkappunt vergelijkbaar zijn op niet-waarneembare karakteristieken (zoals subjectieve kwaliteit, reputatie). Echter, om te onderzoeken of de interpretatie van deze assumptie invloed heeft op de uiteindelijke resultaten zal uiteindelijk een ‘fuzzy’ model (met cross-overs) als basismodel worden gehanteerd, maar in aanvulling daarop een ‘geforceerd sharp’ model waarbij alle cross-overs en no-shows buiten de analyse worden gelaten.

Technisch gezien wordt de ‘fuzzy’-regressiesystematiek geïmplementeerd door in plaats van de waargenomen toekenningsvariabele (de onafhankelijke variabele waarvan de coëfficiënt het effect meet) een instrument te creëren door een schattingsfunctie uit te voeren waarbij toekenning wordt geschat op basis van de juryscore en achtergrondvariabelen via een niet-parametrische probitvergelijking. Dit instrument vervangt vervolgens de onafhankelijke toekenningsvariabele. Helaas is de maximum-likelihood functie voor de OTP-toekenningen niet concaaf¹¹, waardoor het enkel mogelijk is om de ‘fuzzy’ techniek toe te passen voor de thematische programma’s.

3.4.2 *Trendbreuk rondom afkappunt*

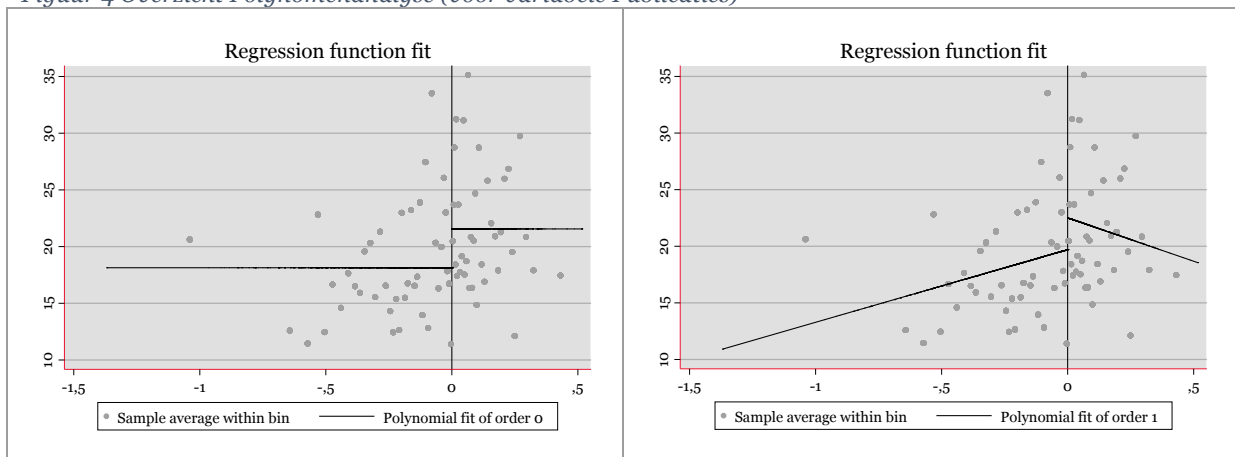
Zoals reeds omschreven is het belangrijkste uitgangspunt van een RDD-analyse dat er geen substantiële verschillen zijn tussen onderzoekers die net boven en net beneden het afkappunt scoorden. Hoewel dit tot op zekere hoogte altijd een assumptie blijft die enkel met logische argumenten met betrekking tot het juryproces kan worden onderbouwd, is het gebruikelijk in ieder geval te controleren of er trendbreuk te constateren is op het afkappunt wanneer we eerdere productiviteit van onderzoekers vergelijken. Deze analyse is uitgevoerd door via t-toets te testen of de groep net boven het afkappunt significant verschilt op productiviteit in de periode vóór indiening met de groep net beneden het afkappunt. Uit deze analyse blijkt dat met bepaalde bandbreedtes (zie sectie 3.3.4) er een lichte trendbreuk is voor de variabelen publicaties en publiek-private publicaties (enkel significant bij $p < 0,1$). Deze constatering is meegenomen in de presentatie van de resultaten.

¹¹ Dit betekent dat de ML-regressie geen (lokaal) minimum vindt, waardoor de geschatte functie geen oplossing heeft. Dit is een veel voorkomend probleem bij niet-parametrische schattingen met een hoge specificatie van het regressiemodel en een relatief beperkt aantal observaties.

3.4.3 Analyse polynomen

Bij een regressie discontinuïteitsanalyse is het gebruikelijk ook een visuele en statistische analyse uit te voeren van het verband dat uiteindelijk geschat wordt in de OLS-vergelijking. Hierbij gaat het met name om de analyse van de relatie tussen de juryscore en de uiteindelijke productiviteitsmaatstaven, zowel boven en onder het afkappunt gezamenlijk als afzonderlijk. Dit wordt meestal gedaan door het presenteren van een ‘regressieplot’, dat via een puntenwolk (waarbij elk punt een aantal observaties vertegenwoordigt) zowel het verband met de juryscore als de discontinuïteit (‘de sprong’) visueel in kaart brengt. Onderstaande figuren geven een voorbeeld van een dergelijke puntenwolk voor de variabele Publicaties (in de implementatieperiode), met zowel geen lineair verband (links) als een lineair verband (rechts). De figuren laten zien dat er duidelijk een ‘sprong’/discontinuïteit is rondom het afkappunt, wat op een effect wijst voor deze maatstaf. Er is echter geen duidelijk effect over de gehele juryscore heen te herkennen, en ook de lineaire verbanden tussen Juryscore en productiviteit zijn niet (of nauwelijks) significant, ook wanneer deze door middel van een restrictie (groter dan 0) in een regressie worden geanalyseerd. Dit betekent in de praktijk dat er geen aantoonbaar verband is tussen een hogere Juryscore en productiviteit. Een uitdaging hierbij is dat veel gangbare RDD-schattingmethoden als uitgangspunt een lineair verband modelleren om de de functie te schatten. Zoals ook zichtbaar is op onderstaand figuur, heeft het toevoegen van een niet-significant lineair verband tussen juryscore en publicaties (of andere outputmaatstaven) een vrij substantieel effect op de schatting van de intercept vanwege het mechanisme achter een OLS-schatting die de totale ‘fit’ van een functie probeert te optimaliseren, en niet alleen een accurate constante term. Gezien het feit dat deze specificatie zo gebruikelijk is en er in bepaalde gevallen een licht significant verband is tussen de juryscore en productiviteit, is gekozen om zowel specificaties zonder lineair verband (de facto het vergelijken van gemiddelden, zoals in de linkerfiguur, zij het wel in een regressie met covariaten), als ook een specificatie mét een lineair verband op te voeren (traditioneel ontwerp).

Figuur 4 Overzicht Polynomenanalyse (voor variabele Publicaties)



3.4.4 Keuze bandbreedte

Aangezien het principe van het RDD-mechanisme conceptueel is gebaseerd op het vergelijken van onderzoekers net boven en onder het afkappunt, is het definiëren van de bandbreedte een belangrijke keuze (wat is “net”). Bij een lineaire schatting met polynomen geldt dit probleem niet, hier is het uitgangspunt dat het inbrengen van de juryscore als covariaat zorgt dat de hele steekproef kan worden meegenomen en men feitelijk de intercept als limiet naar het afkappunt kan interpreteren. Wanneer geen polynomen worden gebruikt en gemiddeldes worden vergeleken is het belangrijk een juiste

bandbreedte te kiezen. Wij gebruiken hiervoor twee gangbare methoden: MSE (Mean Square Error) en CERRD (Cover Error Rate Regressie Discontinuïteit)¹². Beide methoden werken dynamisch, dat wil zeggen dat ze een optimale bandbreedte kiezen die voor elke unieke variabele en specificatie weer opnieuw berekend worden.

3.4.5 Overzicht modelspecificaties

Onderstaand tabel geeft de uiteindelijke verschillende specificaties van het RDD Model weer, met daarin verwerkt de keuze voor de ‘fuzzy’ of ‘sharp’ aanpak, keuze voor bandbreedte specificaties en het meenemen van een polynoom van de juryscore of niet. Zoals vermeldt in sectie 2 werken alle RDD-specificaties slechts met de onderzoekerspopulatie die maximaal 1 honorering heeft ontvangen om overlap te voorkomen. Voor elke analyse zijn als controlevariabelen de variabelen JAAR (als dummy’s) om absolute trends over de jaren heen op te vangen, GESLACHT en en HERINDIENING meegenomen. Deze laatste variabelen controleert of onderzoekers die meerdere malen eenzelfde dossier hebben ingediend een ander effect laten zien. Voor elke specificatie zijn heteroskedastisch-robuste standaardfouten gebruikt, omdat de assumptie van homoskedastisiteit (een constante variatie over verschillende niveaus van productiviteit) zeer waarschijnlijk niet standhoudt voor indicatoren zoals publicaties en citaties.

Tabel 7 RDD specificaties

Indicator	RDD 1	RDD 2	RDD 3	RDD 4	RDD 5	RDD 6
Meenemen meervoudige honoreringen	NEE	NEE	NEE	NEE	NEE	NEE
Meenemen cross-overs & no-shows ('fuzzy' /sharp)	JA	NEE	JA	NEE	JA	NEE
Controlevariabelen	JAAR HERIND ¹³ GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT
Overige specificaties	BB: MSE POL ¹⁴ : 0 ROBUST	BB: MSE POL: 0 ROBUST	BB: CER POL: 0 ROBUST	BB: CER POL 0 ROBUST	BB: Geen POL 1 ROBUST	BB: Geen POL 1 ROBUST

3.5 Specificatie fixed-effects / random-effects (panel) analyse

3.5.1 Fixed-effects vs Random effects

Om met paneldata effecten te schatten zijn er twee hoofdklassen van schattingsfuncties mogelijk: fixed effects (zoals hiervoor besproken) en random effects. In een random effects is het uitgangspunt dat de variatie van de ongemeten variabelen willekeurig is verdeeld over de populatie, terwijl een fixed-effects uitgaat van structurele verschillen tussen typen onderzoekers (bijvoorbeeld die met gehonoreerde en niet-gehonoreerde aanvragen). Omdat de random-effects een efficiëntere en flexiblere methode is heeft deze in principe de voorkeur, mits er dus geen structurele verschillen zijn in de variaties van de ‘error-term’. Dit kan worden gecontroleerd door het uitvoeren van een Hausman test¹⁵. Uit deze

¹² Voor een uitleg van deze methods, zie: Calonica, S; Cattaneo, M, Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. The Stata Journal, VOL. 14, Number 4, pp. 909–946

¹³ Binaire variabele die aangeeft of eenzelfde dossier meerdere malen is ingediend voordat een definitieve beslissing viel

¹⁴ Aantal polynomen van de genormaliseerde juryscore meegenomen in de regressieanalyse

¹⁵ Greene, W. (2011). Econometric Analysis, 7th edition, Pearson, New York.

analyse blijkt dat – zoals ook wel theoretisch verwacht – er inderdaad structurele verschillen zijn tussen onderzoekers en dat hierdoor een fixed-effects aanpak de voorkeur heeft.

Helaas is het problematisch – vanwege zowel praktische als theoretische redenen - om voor niet-parametrische analyses zoals een probit-analyse een fixed-effects aanpak uit te voeren¹⁶ terwijl voor bepaalde maatstaven (binaire en %-variabelen) dat normaal gesproken geadviseerd is. Voor deze variabelen is als terugvaloptie een random-effects aanpak gekozen.

3.5.2 Analyse succesfactoren & Gemeenschappelijke trendanalyse

Een fixed-effects analyse is gebaseerd op het principe van het ‘neutraliseren’ van persoonsgebonden effecten door enkel naar de ontwikkeling over tijd te kijken. Dit gebeurt door het ‘de-meaning’, feitelijk gezien het aftrekken van de gemiddelde productiviteit van een specifieke onderzoeker over de jaren heen van iedere observatie. Technisch gezien kan dit worden geïmplementeerd door het toevoegen van ‘dummy’s’ op onderzoekerniveau in een standaard OLS panel-regressie. Het grote voordeel van een fixed-effects aanpak is dat de statische persoonsgebonden factoren hiermee automatisch geen invloed meer kunnen uitoefenen op de schatting. Dit geldt ook voor niet-gemeten statische karakteristieken. De keerzijde hiervan is dat het overigens onmogelijk is om het effect van statische variabelen (zoals geslacht) te meten. Een fixed-effects aanpak kan dus uitkomst bieden bij een verwachte selectiebias waar niet-gemeten karakteristieken van onderzoekers een verstoring effect kunnen hebben op de schattingsfunctie. Een grote inherente beperking van de fixed-effects aanpak is echter dat deze niet kan controleren voor niet-waarneembare achtergrondkarakteristieken die over tijd veranderen én gerelateerd zijn aan de productiviteitsindicator. Een voorbeeld hierbij zou kunnen zijn een bepaalde mediabekendheid die bijdraagt een succesvolle onderzoekerscarrière, maar niet direct meetbaar is in onze dataset. Dergelijke verschillen hebben geen effect wanneer deze variaties vrij willekeurig over de onderzoekspopulatie zijn verdeeld, maar verstoren de nauwkeurigheid van de schatting wanneer de groep met een toekenning structureel een andere trend vertoont dan de afgewezen groep.

Deze ‘gemeenschappelijke trend’ assumptie kan per definitie niet volledig worden getoetst, omdat een groot aantal waarschijnlijke achtergrondkenmerken niet beschikbaar zijn (bijvoorbeeld iemands motivatie, of persoonlijke netwerk). Wel is het mogelijk te onderzoeken of succesvolle en niet-succesvolle aanvragers significant verschillen qua trend op het gebied van een aantal kernprestatieindicatoren in de periode voor de aanvraag. Hoewel er wel (licht) significante verschillen zijn in trends tussen deze twee groepen, laat de tabel hieronder zien dat deze verschillen in trends geen voorspellende waarde hebben voor honorering of niet. Overigens heeft geen enkele prestatie-indicator een voorspellende werking op selectie, enkel het feit of men eerder een STW-beurs heeft ontvangen heeft een significant effect (elke eerdere honorering geeft gemiddeld 17% meer kans op succes bij een volgende aanvraag). Dit betekent dat er zeker een risico is dat de gemeenschappelijke trendvariabele niet geheel standhoudt, maar dat er ook geen direct bewijs is dat dit direct een grote verstoring te weeg zal brengen. Hierin is de RDD-aanpak complementair, het gebruik maken van het juryselectiemechanisme zorgt dat een gemeenschappelijke trend zeer aannemelijk is voor aanvragers rondom het afkappunt.

Tabel 8 Probit selectiemodel

	<i>OTP (volledig)</i>	<i>TP (volledig)</i>	<i>OTP (max. 1 toekenning)</i>	<i>TP (max. 1 toekenning)</i>
Publicaties (t-1)	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband
Publiek-Private Publicaties (t-1)	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband
Octrooiaanvragen (t-1)	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband

¹⁶ Zie ook Greene (2012) The Bias of the Fixed Effects Estimator in Nonlinear Models.

	<i>OTP (volledig)</i>	<i>TP (volledig)</i>	<i>OTP (max. 1 toekenning)</i>	<i>TP (max. 1 toekenning)</i>
Trend Publicaties (t-3 / t-1)	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband
Trend Publiek-Private Publicaties (t-3 / t-1)	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband
Trend Octrooiaanvragen (t-3 / t-1)	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband
Aantal aanvragen tot die datum	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband
Aantal honoreringen tot die datum	0,17***	Geen sig. verband	NA	NA
Geslacht	Geen sig. verband	Geen sig. verband	Geen sig. verband	Geen sig. verband
JAAR (dummy's)	***	***	Geen sig. verband	Geen sig. verband

3.5.3 Overzicht specifieke specificatie

De tabel hieronder geeft de specificatie weer van de Fixed-en Random Effects modellen. Er is gekozen voor 1 set schattingsfuncties die net als de RDD-analyse enkel onderzoekers meeneemt met maximaal 1 succesvolle aanvraag. Deze specificatie dient vooral als vergelijking voor de RDD-resultaten. In het tweede model zijn OTP en TP-honoreringen als cumulatieve ‘dummy’s’ meegenomen uitgaande van een ‘multiple event’ ontwerp¹⁷. Feitelijk wordt hier cumulatief voor elke honorering een dummy-indicator (per type subsidie) met 1 opgehoogd in de respectievelijke effectperiodes. Wanneer bijvoorbeeld een bepaald jaar voor een bepaalde onderzoeker binnen 3 effectperiodes valt van 3 verschillende beurzen, is de waarde van deze indicator 3. Hiermee worden de effectperiodes als het ware ‘gestapeld’ en is het mogelijk om ook met overlap consistent het effect te schatten. Belangrijk is het hierbij te melden dat dus het geschatte effect het effect is *pér* beurs. Om een eventueel teruglopend marginaal effect van opeenvolgende beurzen mee te kunnen nemen is ook een kwadratische term meegenomen.

Als covariaat zijn jaar-dummy’s meegenomen om te controleren voor algemene trends over de tijd.

Tabel 9 Overzicht modelspecificaties

Indicator	FE 1	RE 1	FE 2¹⁸	RE 2
Meenemen meervoudige honoreringen	NEE	NEE	JA	JA
Kwadratische interactieterm	NEE	NEE	JA	JA
Meenemen cross-overs & no-shows	JA	JA	JA	JA
Controlevariabelen	JAAR	JAAR	JAAR	JAAR
Overige specificaties	ROBUST ¹⁹		ROBUST	

¹⁷ Zie ook Sandler, R. & Sandler, M (2013). Multiple Event Studies in Public Finance and Labor Economics: A Simulation Study with Applications.

¹⁸ Tussen haakjes het kwadratisch effect

¹⁹ Correctie voor heteroskedasticiteit

4 Resultaten

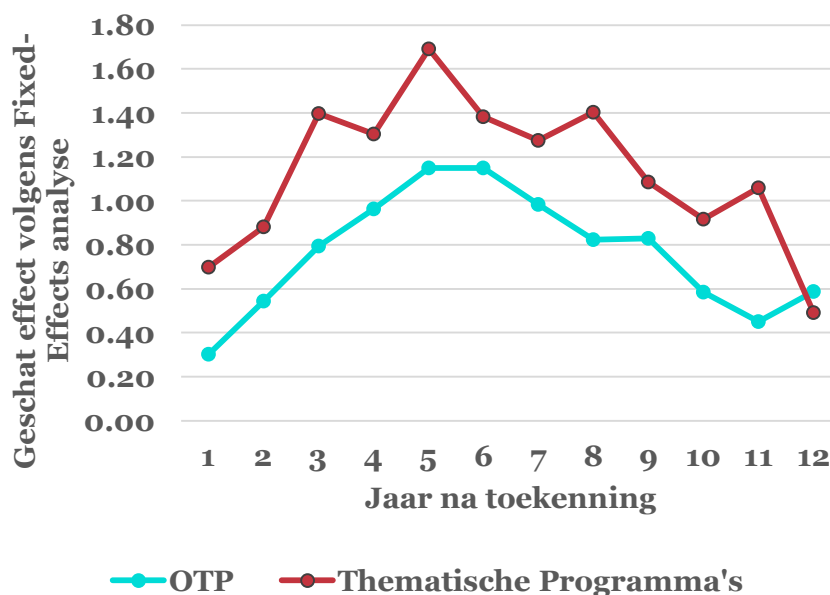
4.1 Overzicht resultaten

De volledige resultaten zijn gepresenteerd op de pagina's hierna, inclusief gedetailleerde methodologische voetnoten. De resultaten zullen besproken en geïntegreerd worden in Sectie 5. Belangrijk is te vermelden dat is gekozen de 'puntschattingen' van de Thematische Programma's voor de RDD-methode niet te presenteren, enkel een t-toets of deze groter zijn dan nul. Dit heeft te maken met het feit dat door een vrij beperkte steekproef (in verband met beperkte beschikbaarheid van jurydata voor Thematische Programma's) en tevens de gevoeligheid van de intercept voor een niet-significante polynoom van de juryscore, een schatting met zeer grote standaardfouten het resultaat was (bijv. 95% confidence-interval tussen de 1 en 48 voor publicaties in de resultaatperiode onder de CERRD methode). Het vermelden van een dergelijke puntschatting, vooral gezien het feit dat deze regelmatig dusdanig buiten de verwachte orde van grootte ligt²⁰, geeft hier een te groot gewicht mee aan deze waarden. Voor OTP is was dit probleem – dankzij een grotere populatie – niet substantieel aanwezig. In het algemeen is gekozen (omdat variatie over het algemeen vrij substantieel is) om in Sectie 5 de confidence intervals te vermelden in plaats van de puntschattingen. Zoals besproken in Sectie 3.4.1 was het helaas niet mogelijk om een fuzzy-regressie uit te voeren voor de OTP-resultaten.

4.2 Aanvullende analyse effectperiode

Om in retrospectief te onderzoeken of de effectperiodes inderdaad daadwerkelijk overeenkomen met de periode waarin een effect waar te nemen is, is een aanvullende analyse over een langere tijdsperiode uitgevoerd via een fixed-effects analyse waarin vertraagde dummy-variabelen van honorering zijn meegenomen. Onderstaande figuur geeft de resultaten van deze analyse voor publicaties weer. Hier is duidelijk te zien dat het effect in de eerste vier jaar na toekenning sterk toeneemt, om te pieken in jaar vijf of zes, waarna het langzaam afneemt (met name vanaf jaar 10). Hieruit blijkt dat de keuze voor een implementatieperiode (jaar 1 tot en met jaar 4) en resultaatperiode (jaar 5 tot en met jaar 8) achteraf gezien vrij nauw samenvalt met de daadwerkelijke effectperiode. Wel is het opvallend dat zelfs 12 jaar na dato er nog een significant positief effect te zien is op onderzoekers die een beurs ontvingen.

Figuur 5 Analyse doorwerking effectperiode



²⁰ De CPB-analyse had vergelijkbare grote variaties. Hier was echter geen significant effect geconstateerd, waarschijnlijk door een nog kleinere populatie.

Tabel 10 Open Technologie Programma's

Indicator	Periode	RDD 1	RDD 2	RDD 3	RDD 4	RDD 5	RDD 6	FE 1	RE 1	FE 2 ²¹	RE 2
Meenemen meervoudige honoreringen		NEE	NEE	NEE	NEE	NEE	NEE	NEE	NEE	JA	JA
Meenemen cross-overs & no-shows		JA	NEE	JA	NEE	JA	NEE	JA	JA	JA	JA
Controlevariabelen		JAAR HERIND ²² GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT	JAAR HERIND GESLACHT	JAAR	JAAR	JAAR	JAAR
Overige specificaties		BB: MSE POL ²³ : 0 ROBUST	BB: MSE POL: 0 ROBUST	BB: CER POL: 0 ROBUST	BB: CER POL: 0 ROBUST	BB: Geen POL 1 ROBUST	BB: Geen POL 1 ROBUST	ROBUST		ROBUST	
Basis N		±300-500 ²⁴	±300-500	±300-500	±300-500	1509	1509	2697	2697	3295	3295
Publicaties	Implementatie	NB(1)	-	NB(1)	-	NB(1)	-	4,16***	NB(3)	4,80*** (-0,17***)	NB(3)
	Resultaat	NB(1)	5,22 (p = 0,142)	NB(1)	-	NB(1)	4,37***	3,99***	NB(3)	4,02*** (-0,13***)	NB(3)
Citatiescore (1)	Implementatie	NB(1)	-	NB(1)	-	NB(1)	-	-	-	-	-
	Resultaat	NB(1)	-	NB(1)	-	NB(1)	-	-	-	-	-
Citatiescore (Leiden Ranking)	Implementatie	NB(1)	-	NB(1)	-	NB(1)	-	-	-	-	-
	Resultaat	NB(1)	-	NB(1)	-	NB(1)	-	-	-	-	-
Aantal publiek-private publicaties	Implementatie	NB(1)	1,16***	NB(1)	0,83**	NB(1)	0,93***	0,81***	NB(3)	1,02*** (-0,03***)	NB(3)
	Resultaat	NB(1)	0,09**	NB(1)	0,88**	NB(1)	1,08***	0,83***	NB(3)	1,06*** (-0,03***)	NB(3)

²¹ Tussen haakjes het kwadratisch effect

²² Binaire variabele die aangeeft of eenzelfde dossier meerdere malen is ingediend voordat een definitieve beslissing viel

²³ Aantal polynomen van de genormaliseerde juryscore meegenomen in de regressieanalyse

²⁴ ± refereert aan het feit dat zowel de MSE als CERRD methode dynamisch een optimale bandbreedte uitrekent die voor elke afhankelijke variabele en specificatie net weer iets anders is.

Indicator	Periode	RDD 1	RDD 2	RDD 3	RDD 4	RDD 5	RDD 6	FE 1	RE 1	FE 2 ²¹	RE 2
Aandeel publiek-private publicaties	Implementatie	NB(1)	-	NB(1)	-	NB(1)	-	-	-	1,18*** (-0,34***)	NB(2)
	Resultaat	NB(1)	-	NB(1)	-	NB(1)	-	-	-	0,84* (-0,29*)	NB(2)
EPO-Octrooiaanvragen	Implementatie	NB(1)	-	NB(1)	-	NB(1)	-	-	-	-	-
	Effect	NB(1)	0,59*	NB(1)	0,57*	NB(1)	0,36***	-	-	-	-
EPO-Patentcitaties	Implementatie	NB(1)	-	NB(1)	-	NB(1)		-	-	-	-
	Resultaat	NB(1)	-	NB(1)	-	NB(1)		-	-	-	-
Succesvolle beursaanvraag Talentprogramma's NWO	Implementatie	NB(1)	NB(2)	NB(1)	NB(2)	NB(1)	NB(2)	NB(3)	-	NB(3)	0,26* (-)
	Resultaat	NB(1)	NB(2)	NB(1)	NB(2)	NB(1)	NB(2)	NB(3)	-	NB(3)	-
Succesvolle beursaanvraag open competities NWO	Implementatie	NB(1)	NB(2)	NB(1)	NB(2)	NB(1)	NB(2)	NB(3)	-	NB(3)	-
	Resultaat	NB(1)	NB(2)	NB(1)	NB(2)	NB(1)	NB(2)	NB(3)	-	NB(3)	0,12*** (-)

NB(1) = Niet beschikbaar i.v.m. onbeschikbare schatting (niet-concaaf). NB(2) Niet beschikbaar i.v.m. te kleine populatie toekenningen; NB(3) Niet beschikbaar i.v.m. onjuist model (niet-parametrisch vereist)

Tabel 11 Thematische Programma's

Indicator	Periode	RDD 1	RDD 2	RDD 3	RDD 4	RDD 5	RDD 6	FE 1	RE 1	FE 3 ²⁵	RE 3
Meenemen meervoudige honoreringen		NEE	NEE	NEE	NEE	NEE	NEE	NEE	NEE	JA	JA
Meenemen cross-overs & no-shows		JA	NEE	JA	NEE	JA	NEE	JA	JA	JA	JA
Overige specificaties		BB: MSE POL: o ROBUST	BB: MSE POL: o ROBUST	BB: CER POL: o ROBUST	BB: CER POL: o ROBUST	BB: - POL 1 ROBUST	BB: - POL 1 ROBUST	ROBUST		ROBUST	
Controlevariabelen		JAAR	JAAR	JAAR	JAAR	JAAR	JAAR	JAAR	JAAR	JAAR	JAAR
N (Basis)		±75	±75	±50	±50	±300	±300	2697	2697	3295	3295
Publicaties	Implementatie	-	-	+* ²⁶	-	-	-	6,96***	NB(2)	6,42*** (-0,22***)	NB(2)
	Effect	+* ²⁷	+***	-	+***	+***	+***	7,36***	NB(2)	7,88*** (-0,28***)	NB(2)
Citatiescore (1)	Implementatie	-	-	-	-	-	-	-	-	-	-
	Effect	-	-	-	-	-	-	-	-	-	-
Citatiescore (Leiden Ranking)	Implementatie	-	-	-	-	-	-	-	-	-	-
	Effect	-	-	-	-	-	-	-	-	-	-
Aantal publiek-private publicaties	Implementatie	-	-	-	-	-	-	0,61***	NB(2)	0,62*** (-0,01**)	NB(2)
	Effect	-	-	-	-	-	-	0,76***	NB(2)	0,80*** (-0,05***)	NB(2)
Aandeel publiek-private publicaties	Implementatie	-	0,08*	0,12*	-	-	-	-	-	-	-

²⁵ Tussen haakjes kwadratisch effect

²⁶ Geconstateerde trendbreuk op eerdere productiviteit (significant tegen *)

²⁷ Vanwege de zeer grote standaardfouten is gekozen hier geen 'puntschatting' weer te geven, maar enkel de resultaten van de t-toets of het effect groter dan 0 is. De grote standaardfouten (bijv. een 95% confidence interval tussen 1 en 39) zijn te wijten aan de substantiele variatie en de kleine steekproef beschikbaar voor de RDD-analyse.

Indicator	Periode	RDD 1	RDD 2	RDD 3	RDD 4	RDD 5	RDD 6	FE 1	RE 1	FE 3 ²⁵	RE 3
	Effect	-	-	-	-	-	-	-	-	-	-
EPO-Octrooiaanvragen	Implementatie	-	-	-	-	-	-	-	-	-	-
	Effect	+**	-	+**	-	-	-	-	-	-	-
EPO-Patentcitaties	Implementatie	-	-	-	-	-	-	-	-	-	-
	Effect	-	-	-	-	-	-	-	-	-	-
Succesvolle beursaanvraag Talentprogramma's NWO	Implementatie	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	-	-	-	-
	Effect	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	-	-		
Succesvolle beursaanvraag open competities NWO	Implementatie	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	NB(2)	-	NB(2)	-
	Effect	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	NB(1)	NB(2)	-	NB(2)	-

NB1 = Niet beschikbaar vanwege te kleine populatie toekenningen. ; NB (2) = Niet beschikbaar i.v.m. niet relevante schattingsmethode (niet-parametrische model vereist)

5 Discussie en Conclusie

Onderstaande tabel vat de belangrijkste uitkomsten van de econometrische effectmeting samen. De indicatoren vergelijken twee verschillende effecten voor elke indicator:

- De periode tijdens implementatie (eerste jaar na aanvraag tot en/met het vierde jaar na aanvraag)
- De periode na implementatie (vijfde jaar na aanvraag tot en/met achtste jaar na aanvraag).

Deze twee tijdsperiodes zijn toegevoegd aan de aanpak uit de CPB-studie, omdat over het algemeen de periode rondom het einde van het project en de periode daarna juist de resultaten zichtbaar worden. De getallen in onderstaande tabel geven de opgetelde effecten weer per periode van vier jaar.

Het is belangrijk te wijzen op de verschillende niveaus van robuustheid van deze resultaten. Onze twee primaire schattingsmethoden; namelijk de RDD-methode en de Fixed/Random Effects analyse (met varianten daaronder) hebben beide verschillende methodologische beperkingen en functioneren daarom aanvullend. De resultaten laten zien dat slechts een beperkt deel van de resultaten consistent te zien zijn in beide schattingsmethoden (groene velden) en dat enkel de indicator aantal publiek-private publicaties indicatoren consistente resultaten laat zien over alle varianten heen. De resultaten van de verschil-in-verschil analyse zijn hieronder niet meegenomen, omdat de Fixed/Random Effects methodes in elk methodologisch opzicht de voorkeur hebben. Het feit dat de resultaten qua het vinden van significante effecten toch enigszins vergelijkbaar zijn versterkt echter wel de algemene betrouwbaarheid van onderstaande bevindingen.

Tabel 12 Overzicht resultaten econometrische effectmeting

WK = Wetenschappelijk; UT = Utiliteit ²⁸	OTP – Tijdens implementatie (t1/t/m t4)	OTP – Na implementatie (t5 t/m t8)	TP – Tijdens implementatie (t1 t/m t4)	TP – Na implementatie (t5 t/m t8)
Publicaties (WK)*	ca. 3-6	ca. 3-5	ca. 5-8	ca. 5-9
Citatie-score (WK)				
Citatie-score (Leiden Ranking) (WK)				
Aantal publiek-private publicaties (UT/WK)*	ca. 0-1	ca. 0-1	ca. 0-1	ca. 0-1
Aandeel publiek-private publicaties (UT)	ca. 0-2%-punt	ca. 0-2%-punt	ca. 1-19%-punt	
EPO-Octrooiaanvragen (UT)*		ca. 0-1		+ ²⁹
EPO-Octrooicitaties (UT)				
Succesvolle beursaanvraag Talentprogramma's NWO (WK)	ca. 1-50%-punt			
Succesvolle beursaanvraag open competities NWO (WK)		3-32%-punt		

Samenvatting Resultaten Econometrische Analyse. Groen = consistente resultaten in beide toegepaste modellen (RDD en FE/RE) Grijs: Significante bevindingen in één van de twee hoofdmodellen (RDD of FE/RE). Gegeven intervallen zijn hier de breedste range van de fixed effects schattingsmethodiek, afgerond naar hele eenheden.

²⁸ * ook uitgevoerd in CPB analyse in 2013

²⁹ Zie Sectie 4.1 voor een discussie waarom hier geen puntschatting is vermeld.

De volgende resultaten vallen op:

- Deelname aan **Thematische Programma's lijkt een positief effect te hebben op het aantal wetenschappelijke publicaties**, in de orde grootte van 5-8 publicaties tijdens implementatie van het project en tussen de 5-9 publicaties in de periode rondom het einde van het project. Deze uitkomst komt vrij nauw overeen met de CPB-bevindingen. Wel is het waarschijnlijk dat deze schattingen eerder aan de hoge dan aan de lage kant zijn, omdat er qua publicatiegedrag niet alleen absolute verschillen zijn op het moment van aanvraag (waar voor wordt gecorrigeerd), maar er zou ook sprake kunnen zijn van een trendverschil op niet-waarneembare factoren, zoals reputatie of netwerk. Dit laatste zou eventueel kunnen leiden tot overschattingen van de effecten.
- In tegenstelling tot de CPB-analyse vindt deze schatting een grote **waarschijnlijkheid van ook een positief effect op het aantal publicaties in de periode rondom het einde van het project voor OTP-beurzen**, in de orde grootte van 3 tot 5 publicaties. Hiervoor geldt dezelfde beperking als voor de TP-beurzen. Voor effect gedurende de implementatie is geen consistent effect te vinden.
- In termen van **aantal publiek-private publicaties³⁰ is er een consistent effect van ongeveer 1 publicatie voor OTP en eenzelfde aantal maar dan niet consistent voor Thematische Programma's**. De CPB-studie stelde een vergelijkbaar effect vast, maar enkel voor TP-beurzen. Vanwege de correlatie met het totaal aantal publicaties is een aanvullende analyse gedaan voor het aandeel van publiek-private publicaties in het totaal aantal publicaties (per periode). Hier zijn de resultaten niet consistent tussen de modellen, maar wel allemaal positief. Er is **al met al dus gedeeltelijk bewijs voor een additioneel effect van STW-beurzen op de mate van co-publicaties met bedrijven**, met name omdat succesvolle aanvragers dit reeds al meer deden en omdat deze vorm van co-publiceren relatief weinig voor komt.
- Er is **geen effect van STW-beurzen op de kwaliteit van de wetenschappelijke output** gemeten als gewogen citatiescore. Variaties hierin zijn vrij beperkt wanneer men kijkt naar portfolio's over vier jaar hier.
- **Octrooiaanvragen (en Octrooicitaties) lijken over het algemeen niet of nauwelijks aantoonbaar te worden beïnvloed** door toekenning van een STW-beurs. Er is enkel beperkt en inconsistent bewijs voor een kleine stijging van octrooiaanvragen in de periode. In tegenstelling tot de CPB-analyse vinden we echter ook geen negatief effect.
- Er is slechts inconsistent en beperkt bewijs dat deelname aan OTP-programma's een positieve invloed heeft op de kans om een andere NWO-beurs te krijgen. De geringe aantallen maken het lastig om een nauwkeurige schatting te maken.

³⁰ Gedefinieerd als publicaties met een coauteur waarvan een bedrijfsnaam in de adresgegevens staat.

technopolis |group| The Netherlands
Spuistraat 283
1012 VR Amsterdam
The Netherlands
T +31 20 535 2244
F +31 20 428 9656
E info.nl@technopolis-group.com
www.technopolis-group.com