

DETERMINANTEN VAN ONDERZOEKSOUTPUT

Het gebruik van publicaties en citaties als instrumenten voor het beoordelen van academici – denk aan promotiebeslissingen – is het afgelopen decennium sterk toegenomen, ook in Europa. Ook de toekenning van onderzoeksfondsen gebeurt steeds meer op basis van excellentiecriteria wat de concentratie van onderzoeksgelden in de handen van topwetenschappers in de hand werkt. Aansluitend bij het toenemende belang van kwantitatieve outputmeting neemt ook de aandacht van economen voor het verklaren van academische onderzoeksproductiviteit toe (zie o.a. Stephan & Levin (1992) en Stephan (1996) voor een overzicht). Het bestaande empirische onderzoek wordt echter gekenmerkt door een focus op het verklaren van gemiddelde onderzoeksproductiviteit. Dit is een heel restrictieve invalshoek op het fenomeen onderzoeksproductiviteit, gegeven dat de productiviteitsverdeling typisch heel scheef is. Figuur 1 bevestigt deze scheve verdeling van onderzoekers (ZAP) in biomedische en exacte wetenschappen aan de K.U.Leuven voor de periode 1992-2001, en dit voor zowel publicatie- als (en vooral voor) citatieoutput¹. Op de horizontale as worden individuele onderzoekers geordend volgens hun onderzoeksoutput d.w.z. van minder tot meer productief. De zogenaamde Lorenzcurve geeft het verband tussen een bepaald deel van de onderzoekerspopulatie en het overeenstemmende aandeel in de totale onderzoeksoutput. De figuur toont dat de 25% meest productieve onderzoekers verantwoordelijk zijn voor meer dan 75% van het aantal publicaties en tot 90% van het totale aantal citaties.

Onze analyse van de determinanten van publicatieoutput houdt rekening met deze grote verscheidenheid tussen onderzoekers. In een eerste luik kijken we naar determinanten van top publicatieoutput. We stellen ons de vraag wat iemand tot een toponderzoeker maakt? En eens men topstatus bereikt heeft, welke factoren zorgen dan voor een persistent hoge productiviteit? Onze aandacht gaat hierbij voornamelijk uit naar het contrasteren van enerzijds exogene factoren, zoals leeftijd en geslacht, en anderzijds institutionele factoren, zoals het onderzoeksklimaat, het promotiebeleid, het onderzoeksfinancieringssysteem en dergelijke. In een tweede luik verruimen we ons blikveld en gaan we na in welke mate het effect van determinanten van onderzoeksproductiviteit varieert langsheen de gehele outputverdeling. Bijvoorbeeld: speelt het in de literatuur uitgebreid gedocumenteerde 'gender effect' (waarbij de gemiddelde mannelijke onderzoeker productiever blijkt dan de gemiddelde vrouwelijke onderzoeker) ook een belangrijke rol voor toponderzoekers of treedt dit effect voornamelijk op voor minder productieve onderzoekers?

TOP WORDEN EN TOP BLIJVEN

In onze analyse van top publicatieoutput voeren we eerst een clustering uit van het aantal publicaties, controlerend voor wetenschappelijke discipline en tijdstrend. Dit laat ons toe om ieder jaar de toponderzoekers te scheiden van de gemiddelde en de weinig productieve wetenschappers. Ongeveer een kwart van de onderzoekers in de dataset behoort in de observatieperiode minstens éénmaal tot de

Top worden

Om te analyseren welke factoren een rol spelen in het proces naar topperformantie schatten we een Cox model² met de duurtijd tot de eerste top output van de onderzoeker als afhankelijke variabele (zie het eerste model in Tabel 1). Een eerste resultaat bevestigt het meer algemene 'gender effect' dat vaak gerapporteerd wordt in analyses van publicatieoutput: de kans op topperformantie is typisch kleiner voor vrouwelijke wetenschappers. Daar waar leeftijd geen merkbaar effect heeft op de waarschijnlijkheid om tot de meest productieve onderzoekers te behoren, spelen institutionele factoren wel een rol: departementshoofden en (gewoon) hoogleraren maken meer kans om 'top' te zijn inzake publicatieoutput dan onderzoekers in lagere hiërarchische posities. Daarnaast valt ook de rol van onderzoeksfinanciering op: wetenschappers die toegang hebben tot substantiële bijkomende middelen onder de vorm van GOA-financiering maken meer kans om topstatus te bereiken⁴.

Ten slotte vinden we slechts beperkte steun voor de stelling dat bijkomende onderwijsbelasting zou leiden tot verminderde kansen om een top onderzoeksoutput te behalen wegens een substitutie-effect.

Top blijven

Teneinde de persistentie in topprestaties verder te analyseren, schatten we een Cox model met de duurtijd tot de eerste of volgende topprestatie als afhankelijke variabele

Publiceren of Sterven? Inzichten in

Stijn Kelchtermans en Reinhilde Veuglers

Deze grote ongelijkheid in onderzoeksoutput suggereert een aanzienlijke heterogeniteit tussen academische onderzoekers. Enkel aandacht besteden aan het gemiddelde aantal publicaties en citaties vertelt dan ook slechts een deel van het verhaal.

topcategorie. Zes op honderd wetenschappers is persistent top d.w.z. behoort iedere periode² tot de topcategorie. Onderzoekers zijn erg persistent: degenen met een hoge output in één periode hebben een grote kans om deze prestatie de daaropvolgende periode te herhalen. Een analoge observatie is van toepassing voor onderzoekers met een lage output.

(zie het tweede model in Tabel 1). De resultaten leveren de volgende belangrijke inzichten op.

Vooreerst vinden we overtuigend bewijs voor het zogenaamde accumulatief karakter van topprestaties: voor iedere topprestatie uit het verleden neemt de kans van de onderzoeker om dat outputniveau te handhaven met ongeveer een derde toe. Opmerkelijk is dat dit fenomeen nog sterker van toepassing is op vrouwen dan op mannen: voor vrouwelijke onderzoekers is de kans om opnieuw tot het koppeloton te behoren meer dan dubbel zo hoog voor iedere keer dat ze top waren in het verleden. Dit suggereert dat het gender effect in de eerste plaats moet gezien worden als een selectieprobleem ten aanzien van het bereiken van top status, eerder dan dat het een cruciale rol zou blijven spelen voor daaropvolgende topprestaties. Daar waar extra onderzoeksfinanciering en het voorzitterschap van een departement van belang bleken voor het verklaren van het bereiken van een eerste topstatus zijn ze minder van betekenis voor het voorspellen van herhaalde topprestaties.

De correctie voor niet-geobserveerde individuele verschillen tussen onderzoekers in dit model is significant, wat suggereert dat factoren zoals onderzoekstalent integraal deel uitmaken van de verklaring van topperformantie en de persistentie ervan.

DE REPRESENTATIEVE ONDERZOEKER ... BESTAAT NIET

Na deze analyse van 'star scientists' verruimen we ons blikveld. We stellen ons de vraag of de determinanten van publicatieoutput zoals die hierboven aan bod kwamen in dezelfde mate van belang zijn voor de lagere output-

TABEL 1: COX MODELLEN VOOR TOP PUBLICATIEOUTPUT

Variabele	Duurtijd tot eerste top output		Duurtijd tot eerste of volgende top output	
	Hazard ratio	z	Hazard ratio	z
man	1.86**	(2.06)	2.32***	(2.88)
leeftijd	0.91	(-0.89)	0.96	(-0.47)
leeftijd ²	1.00	(0.14)	1.00	(-0.35)
graad♣				
docent in t-1	0.13***	(-6.45)	0.18***	(-6.97)
hoofddocent in t-1	0.24***	(-6.92)	0.35***	(-5.88)
hoogleraar in t-1	0.36***	(-4.41)	0.52***	(-4.40)
departementshoofd in t-1	1.52**	(2.51)	1.15	(1.35)
onderzoeksfinanciering				
(co-)promotor van GOA-project in t-1	1.77***	(2.78)	1.25	(1.63)
(co-)promotor van OT-project in t-1	0.92	(-0.36)	1.07	(0.57)
onderwijsbelasting (jaaruren)	0.96*	(-1.82)	0.98	(-1.41)
aantal voorgaande top outputs			2.33***	(4.19)
man * aantal voorgaande top outputs			0.56***	(-2.89)
Observaties		4,423		5,562

Het Cox model voor herhaalde top output controleert voor (niet-geobserveerde) individuele heterogeniteit d.m.v. een random effect. Opgenomen controlevariabelen: hoofd discipline van de onderzoeker, faculteit, anciënniteit als ZAP en in huidige graad, voltijds aan de universiteit, benoeming als ZAP in of na 1992.

♣ basiscategorie = gewoon hoogleraar

* p<0.10, ** p<0.05, *** p<0.01

¹ De publicatie- en citatiegegevens zijn gebaseerd op de ISI Science Citation Expanded Index (SCIE).

² Hiervoor hanteren we een voortschrijdend venster van twee jaar om rekening te houden met 'publication lags'.

³ De resultaten zijn gerapporteerd als het effect op de 'hazard ratio' d.w.z. de kans om voor de n-de maal top te zijn, gegeven dat men n-1 maal top geweest is.

⁴ Voorzichtigheid is geboden bij de interpretatie van deze resultaten: hoewel het model potentieel endogene determinanten opneemt als vertraagde ('lagged') variabelen kan een omgekeerde causaliteit niet uitgesloten worden.

niveaus. Er zijn grosso modo twee redenen aan te halen waarom men een verschillend effect van determinanten van onderzoeksoutput zou verwachten voor minder versus meer productieve onderzoekers.

Ten eerste is het mogelijk dat de intensiteit van de prikkel om te publiceren voor een gegeven determinant verschillend is naargelang de productiviteit van de onderzoeker. Zo postuleren we dat in de rechterstaart van de productiviteitsverdeling, dus voor de zeer productieve onderzoekers, de onderzoeksoutput hoofdzakelijk bepaald wordt door (niet rechtstreeks geobserveerd) talent en minder door externe 'incentives' zoals bijkomende onderzoeksfinanciering. De intuïtie is dat in de hogere regionen van de productiviteitsverdeling het onderzoekstalent de andere determinanten gaat overheersen én dat het effect van de prikkel van externe incentives en geobserveerde kenmerken (leeftijd, geslacht ...) afneemt met productiviteit.

Ten tweede kunnen zeer productieve onderzoekers verwacht worden onderhevig te zijn aan een verschillende mate van substitutie van onderzoek door andere activiteiten dan minder productieve onderzoekers. Beschouw andermaal het 'gender effect'. Als mogelijke verklaring wordt vaak aangehaald dat de (wederom, gemiddelde) vrouwelijke onderzoeker voor een grotere uitdaging staat om haar onderzoekstijd te vrijwaren dan de gemiddelde mannelijke onderzoeker wegens haar grotere rol inzake gezinsverplichtingen. Als we nu weer de productiviteit van de onderzoeker in aanmerking nemen als conditionerende

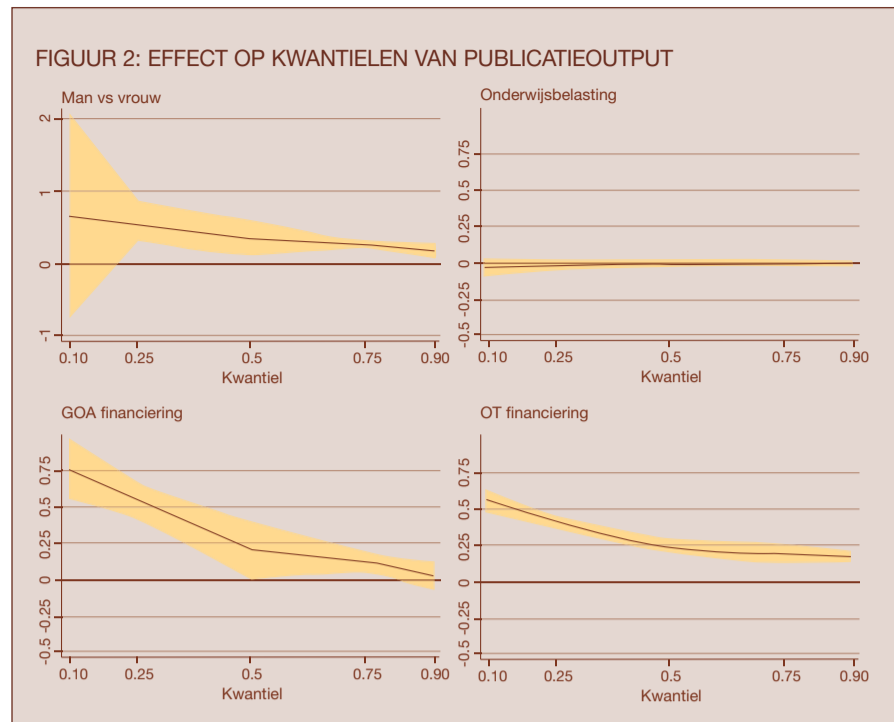
name vooral de output van de minder productieve onderzoekers verhoogt.

Ten slotte vermelden we dat het effect van de beschouwde variabelen op publicatieoutput statistisch verschilt overheen de vijf kwantielen. Dit bevestigt de idee dat men tot een volledig beeld komt van publicatiegedrag door het analyseren van verschillende punten uit de scheve distributie.

CONCLUSIES

Het nog steeds toenemende belang van kwantitatieve indicatoren van onderzoeksoutput in de academische wereld vraagt om een grondig inzicht in de determinanten van onderzoeksproductiviteit.

Onze analyse van top onderzoeksoutput maakt duidelijk welke individuele kenmerken en beleidsvariabelen een impact hebben en welke verschillen er bestaan in het bereiken versus het bestendigen van top performantie. Daarnaast tonen de resultaten aan dat een meer genuanceerde analyse van onderzoeksproductiviteit - die verder



Met betrekking tot het toekennen van onderzoeksfinanciering geven de resultaten aan dat dit aanleiding zal geven tot een reductie in de ongelijkheid inzake publicatieoutput: het positieve effect is groter in de lagere kwantielen dan aan de top van de verdeling. Dit betekent niet dat de minder productieve onderzoekers beter in staat zouden zijn onderzoeksfondsen te valoriseren onder de vorm van publicaties dan de heel productieve onderzoekers. De schattingen zijn immers conditioneel op de actuele

de onderzoeksproductiviteit van academici

factor, kan men postuleren dat vrouwelijke onderzoekers die veel publiceren minder onderhevig zijn aan het gender effect dan hun minder productieve collega's omdat zij via hun talent beter verschillende activiteiten in complement kunnen uitvoeren.

Om de impact van de determinanten op verschillende plaatsen in de publicatiedistributie te bepalen hanteren we kwantielregressies: we bepalen het effect van - bijvoorbeeld - geslacht op het 10de, 25ste, 50ste, 75ste en 90ste kwantiel. We onderwerpen de publicaties hierbij eerst aan een randomisering om kwantiel-schatting mogelijk te maken (Machado & Santos Silva, 2005). Om te corrigeren voor niet-geobserveerde verschillen tussen onderzoekers schatten we een 'correlated random effects' model (Chamberlain, 1982), gebruik makend van de tijdsdimensie in de data.

Figuur 2 toont de resultaten⁵ voor enkele van de variabelen die in de analyse van top publicatieoutput ook al aan bod kwamen. Het gender effect vertoont een opvallend patroon: hoewel er aan de onderkant van de verdeling geen significant verschil gevonden wordt tussen beide geslachten, is dit wel het geval voor de meer productieve wetenschappers. We zien een positief effect in het voordeel van de mannelijke onderzoekers dat echter afneemt en zeer klein wordt aan de top van de verdeling. Het negatieve substitutie-effect van onderwijs op onderzoek bestaat, maar enkel voor de minder productieve onderzoekers en bovendien is het effect erg klein. Het effect van bijkomende onderzoeksfinanciering (gemeten als (co-)promotor van een GOA- of OT-project) is significant positief. Interessant om vast te stellen is dat het effect van deze incentive afneemt in de hogere regionen van de distributie en dus met

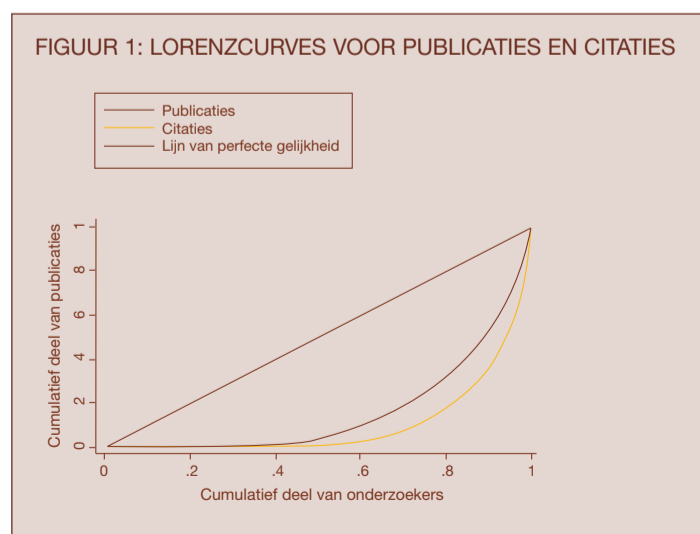
gaat dan een focus op het verklaren van het gemiddelde - nodig is als men het fenomeen ten volle wil begrijpen. Meer in het bijzonder bevestigen de resultaten de hypothese dat de top van de verdeling hoofdzakelijk gestuurd wordt door onderzoekstalent en/of door een gradueel verminderend stimulerend effect van incentives zoals promotiekansen of onderzoeksbeurzen. Het effect van dergelijke incentives is wél waarneembaar in de lagere kwantielen van de verdeling.

Hoewel voorzichtigheid geboden is bij het veralgemenen van de resultaten gebaseerd op de analyse van één instelling, zijn ze informatief met betrekking tot het management van wetenschappers. In het bijzonder bieden de bevindingen concrete aanknopingspunten voor een goed onderbouwde beleidsvoering inzake stimulanprogramma's en mechanismen voor het toekennen van onderzoeksfinanciering. Zo suggereren de resultaten bijvoorbeeld dat een gereduceerde onderwijsbelasting voor minder productieve wetenschappers met als doel hen 'aan boord te nemen', een minder efficiënt instrument is om publicaties te stimuleren.

verdeling van onderzoeksfondsen die ongelijk is aangezien de toponderzoekers de grootste financiering ontvangen. Desalniettemin suggereren de resultaten afnemende meeropbrengsten en kunnen bijgevolg als waarschuwing gezien worden tegen een al te extreme selectiviteit in het toekennen van onderzoeksfondsen.

REFERENTIES

- Chamberlain, G. 1982. Multivariate regression models for panel data. *Journal of Econometrics*, 18: 5-46.
- Machado, J.A.F. & Santos Silva, J.M.C. 2005. Quantiles for counts. *Journal of the American Statistical Association*, 100(472): 1226-1237.
- Stephan, P. E. 1996. *The Economics of Science*. *Journal of Economic Literature*, 34(3): 1199-1235.
- Stephan, P. E. & Levin, S. G. 1992. *Striking the mother lode in science: The importance of age, place, and time*. New York: Oxford University Press.



Stijn Kelchtermans

is als docent verbonden aan de Hogeschool-Universiteit Brussel en is geaffilieerd onderzoeker aan de onderzoekseenheid bedrijfseconomie, strategie en innovatie (MSI)
stijn.kelchtermans@econ.kuleuven.be



Reinhilde Veugelaers

is als gewoon hoogleraar verbonden aan de onderzoekseenheid bedrijfseconomie, strategie en innovatie (MSI)
reinhilde.veugelaers@econ.kuleuven.be



⁵ De solide curves op de grafieken tonen de parameterwaarden. Deze zijn te interpreteren als het marginale effect op het betreffende kwantiel. De gestreepte curves markeren het 95% betrouwbaarheidsinterval van de parameterschattingen.