

Kwaliteit en productiviteit van economisch onderzoek

Tom Groot

SAMENVATTING In 2008 en 2009 zijn de onderzoeksprogramma's van economie, bedrijfseconomie en bedrijfskunde voor de derde keer aan een formele beoordeling onderworpen. Inmiddels is kwaliteitsbeoordeling van onderzoek een algemeen geaccepteerd en belangrijk onderwerp geworden. Een goed kwaliteitsoordeel is in toenemende mate belangrijk voor het verkrijgen van middelen, voor het aantrekken van goede onderzoekers en voor de academische reputatie van de onderzoeksgroep. Het kwaliteitsoordeel kan tot stand komen op basis van oordelen van vakgenoten, de zogenaamde *peer review*, en op basis van *bibliometrische informatie*, zoals het aantal publicaties en citaties. Dit artikel analyseert in hoeverre beide beoordelingsmethoden tot dezelfde conclusies leiden. Daarbij blijken internationale A- en B-publicaties veruit de best verklarende variabelen te zijn. Vervolgens wordt geprobeerd om vast te stellen welke organisatiekenmerken van onderzoeksgroepen verantwoordelijk zijn voor verschillen in kwaliteit en productiviteit van onderzoekoutput. Omvang van de groep blijkt positief samen te hangen met kwaliteit maar negatief met productiviteit.

RELEVANTIE VOOR DE PRAKTIJK Effectief management van onderzoek is belangrijk voor de academische positie van het Nederlandse onderzoek. Daarbij hoort in eerste instantie het kiezen van een goede wijze van onderzoeksbeoordeling, want deze bepaalt de spelregels waar onderzoekers zich aan houden en waardoor ze zich gemotiveerd voelen. Hoe groter de voorspellende waarde van de beoordelingsmethode in universiteiten voor de internationale visitatie is, des te effectiever het interne management is. Uit dit onderzoek blijkt ook dat er objectieve omstandigheden zijn die de productiviteit en kwaliteit van onderzoek beïnvloeden. Ook deze factoren kunnen door het universitaire management doelgericht worden beïnvloed, waardoor de kansen van het Nederlands onderzoek op een goede internationale beoordeling kunnen worden verbeterd.

1. Inleiding

Onderzoekers en onderzoeksprogramma's worden steeds meer op basis van kwaliteit en kwantiteit van onderzoeksprestaties beoordeeld. Zo worden in de VS vaste benoemingen en promotiebeslissingen van onderzoekers

bepaald door hun onderzoeksprestaties (Ballas & Theoharakis, 2003). In het Verenigd Koninkrijk wordt de financiering van onderzoeksgroepen bepaald door de periodieke *Research Assessment Exercise* (Whittington, 1993, 1997). Ook in landen als Spanje, Finland en Nederland vinden periodieke onderzoeksevaluaties plaats (Loft, Jorissen, & Walton, 2002). Over het algemeen worden in deze beoordelingen twee benaderingen gevolgd: de *peer review* en de *bibliometrische methode*. De beoordeling is bij *peer review* gebaseerd op een oordeel van vakgenoten, bij de *bibliometrische methode* is die gebaseerd op informatie over publicaties en citaties. Beoordeling van onderzoeksoutput is in Nederland een relatief recent verschijnsel. Tot 1983 was de financiering van onderzoek direct afhankelijk van het aantal studenten (Groot, 1999). In 1983 is een systeem van *voorwaardelijke financiering van onderzoek* ingevoerd. In de eerste tien jaar werden onderzoeksgelden voornamelijk toegekend op basis van goedgekeurde voorstellen van wetenschappelijke onderzoeksprogramma's. Dat heeft tot enige inzichtelijkheid van onderzoeksinspanningen geleid, maar niet tot stimulering van onderzoeksprestaties. In 1993 is een systeem van periodieke beoordeling van onderzoeksprestaties van start gegaan dat onder toezicht van de Vereniging van Samenwerkende Nederlandse Universiteiten (VSNU) wordt uitgevoerd. Het gaat hier derhalve om een systeem waarin de Nederlandse universiteiten zelf de beoordeling van het onderzoek organiseren. Deze beoordelingen hebben geen directe gevolgen voor de overheidsbekostiging van het onderzoek, maar zijn gericht op drie doelstellingen: (1) het bevorderen van de kwaliteit van het onderzoek, (2) het informeren van faculteitsbesturen over de resultaten die onderzoeksgroepen behalen en (3) het afleggen van verantwoording aan het publiek over de wijze waarop onderzoeksgelden worden besteed (VSNU, 1994; Westerheijden, 1996; VSNU, 2002). De VSNU maakt voor de beoordelingen gebruik van het *peer review*-systeem door een internationale evaluatiecommissie te benoemen die de prestaties van onderzoeksprogramma's beoordeelt. In toenemende mate maken faculteiten inmid-

dels gebruik van bibliometrische informatie om tussen-tijds de onderzoeksprestaties in hun faculteiten te beoordelen en te belonen. Beide methoden werken verschillend en hebben ook zo hun eigen voor- en nadelen.

2. Peer review en bibliometrische methode

Beoordelingen van onderzoek op basis van peer review komen tot stand door percepties van goed geïnformeerde deskundigen. Zij zijn in staat om verschillende kwaliteitsdimensies tegelijkertijd te beoordelen, zij kunnen bij de beoordeling rekening houden met specifieke omstandigheden waaronder de prestaties tot stand zijn gekomen en met de toekomstige mogelijkheden van de onderzoekers. Daar staat echter tegenover dat hun beoordelingen inherent subjectief zijn: ze zijn daarmee ook afhankelijk van subjectieve voorkeuren, de ervaring en de kennis van de beoordelaars (Nederhof & van Raan, 1987; Brinn, Jones, & Pendlebury, 1996). Als de beoordeling wordt uitgevoerd door een commissie is het oordeel van die commissie daarmee ook medeafhankelijk van haar samenstelling. De bibliometrische methode is bedoeld om meer objectiviteit in de beoordeling te brengen. Maar ook deze aanpak heeft zijn tekortkomingen. Zo is de beoordeling nu afhankelijk geworden van de gekozen meetmethode en moet deze ook worden aangepast aan publicatie- en citatiegewoonten in verschillende wetenschapsgebieden (Nederhof & van Raan, 1993). De meest eenvoudige methode is het tellen van het aantal publicaties, maar deze aanpak gaat voorbij aan verschillen in wetenschappelijke kwaliteit (Dwyer, 1994; Hasselback, Reinstein, & Schwan, 2000). Dit kan worden gecompenseerd door te kijken naar het aantal citaties dat naar de beoordeelde publicatie verwijst. Maar ook dat heeft zo zijn tekortkomingen. De uitkomst wordt nu medebepaald door het aantal bronnen (meestal tijdschriften) dat in de analyse wordt betrokken. Een belangrijke bron is het Institute of Scientific Information (ISI) van Thomson Reuters dat een database onderhoudt waarin citaties van duizenden wetenschappelijke tijdschriften worden vastgelegd. Het ISI berekent verschillende citatie-indexen, waarvan de SSCI (Social Science Citation Index) voor in dit artikel onderzochte disciplines relevant is. De SSCI baseert zich op ongeveer 2500 tijdschriften in de sociale wetenschappen en berekent op basis van citaties de impactscore van elk van deze tijdschriften. Helaas zijn lang niet alle tijdschriften in deze systematiek meegenomen. Zo ontbreken veel van de accountingtijdschriften in de ISI-database. Daarnaast is het aantal citaties niet altijd een betrouwbare indicator van wetenschappelijke kwaliteit. Citaties kunnen ook gebruikt worden om wetenschappelijk werk te bekritisieren. Ook kan het aantal citaties kunstmatig hoog worden als auteurs naar zichzelf verwijzen (autocitaties), naar bekende auteurs, tijdschriften of naar tijdelijk populaire onderwerpen. Reviewartikelen en

methodologische papers (Nederhof & van Raan, 1987, 1993) genereren over het algemeen een meer dan gemiddeld aantal citaties. Een andere manier om kwaliteit in de meting in te brengen is om de kwaliteit van het tijdschrift in de beoordeling te betrekken door de (gecorrigeerde) impactscore van het tijdschrift te berekenen. (Newman & Cooper, 1993; van Fleet, Mc-Williams, & Siegel, 2000). Een impactscore geeft aan hoe vaak een gemiddeld artikel in dat tijdschrift is geciteerd door andere tijdschriften. Hierbij echter gelden dezelfde beperkingen als bij artikelsgewijze citatiescores. Het is niet duidelijk in hoeverre de peer review en de bibliometrische methode tot dezelfde uitkomsten leiden. Singleton (1996) vond een lage correlatie tussen subjectieve en citatiegebaseerde beoordelingen. Nederhof en Van Raan (1987, 1993) concluderen op basis van twee Nederlandse studies dat de beide soorten beoordelingen elkaar aanvullen, terwijl Rinia tot de conclusie komt dat er geen relatie bestaat tussen peer review-resultaten en het totaal aantal publicaties (Rinia, et al., 1998).

3. Conditie voor goed onderzoek

Naast de vraag naar de beoordeling kan men zich ook afvragen onder welke condities onderzoek tot bloei kan komen. Sommige studies laten zien dat de productiviteit van onderzoek positief wordt beïnvloed door omvang van de onderzoeksgroep en de omvang van externe financiering. Overigens heeft externe financiering twee gezichten: onderzoek voor commerciële opdrachtgevers blijkt negatief samen te hangen met wetenschappelijke productiviteit, terwijl externe financiering door een wetenschappelijk onderzoeksfonds (denk aan de Nederlandse Organisatie voor Wetenschappelijk Onderzoek NWO en aan Europese onderzoeksfondsen) de wetenschappelijke productiviteit juist bevordert (Johnes, 1988; Cherchye & van-den Abeele, 2005). Ik verwacht derhalve dat de omvang van de onderzoeksgroep, het aandeel tweedegeldstroomfinanciering en daarvan afgeleid het aantal promovendi positief zal samenhangen met de kwaliteit en productiviteit van het onderzoeksprogramma. Daarnaast verwacht ik dat de aard van de discipline een invloed zal hebben op de productiviteit. Bedrijfseconomie en bedrijfskunde worden over het algemeen beschouwd als disciplines met een grote diversiteit aan benaderingen en een niet duidelijk ontwikkeld wetenschapsparadigma, zeker in vergelijking met economie, econometrie en operations research (Kuhn, 1970; Pfeffer, 1993). Een sterk ontwikkeld kennisgebied, theoretische benadering en methodologische aanpak verhoogt de kansen op externe financiering (Lodahl & Gordon, 1973), heeft een grotere onderzoeksproductiviteit (Konrad & Pfeffer, 1990), hebben een kortere reviewtijd van publicaties (Beyer, 1978) en worden meer geciteerd door andere disciplines (Baron & Hannan, 1993; Pfeffer, 1993; Sterken &

Leeflang, 2000). Ik veronderstel derhalve dat onderzoeksgroepen in bedrijfskunde en bedrijfseconomie minder productief zijn dan onderzoekers in de algemene economie en dat deze weer minder productief zijn dan onderzoekers in de econometrie en operations research.

4. Twee externe visitaties van het Economie onderzoek

De evaluatie van 1995

De eerste landelijke onderzoeksvisitatie in Economie onder auspiciën van de VSNU vond plaats in 1995 en werd uitgevoerd door een commissie van zes internationale deskundigen. Zij beoordeelden de resultaten van 92 onderzoeksprogramma's over de periode 1990 tot 1994. De beoordeling vond plaats op basis van informatie die door de programma's werd verstrekt volgens een door de VSNU bepaalde structuur (VSNU, 1994). De gebruikte middelen, de inputs, werden gerapporteerd in aantal fulltime equivalenten personeel, onderverdeeld naar aard (PhD's en senior onderzoekers) en naar herkomst van financiering. Hierbij onderscheidt men drie geldstromen: de eerste geldstroom is afkomstig van de eigen universiteit, de tweede geldstroom komt van landelijke onderzoeksfinanciering door NWO en de derde geldstroom betreft inkomsten van contractonderzoek, gefinancierd door externe, meest commerciële opdrachtgevers. De productie werd verantwoord naar aantallen publicaties in verschillende rubrieken: dissertaties, artikelen in wetenschappelijke tijdschriften, boeken, boekbijdragen (hoofdstukken in boeken) en professionele publicaties. Alle publicaties werden voorzien van een aanduiding van wetenschappelijke kwaliteit door toepassing van een score behorend bij de kwaliteit van een tijdschrift of uitgever (A is de hoogste kwaliteit, F de laagste). Deze score was ontwikkeld door de vaste commissie voor wetenschapsbeoefening van de economische faculteiten van de VSNU en bevat een lijst van 1250

tijdschriften (VSNU-VVCW, 1991). Naast deze informatie ontving de commissie ook vijf toppublicaties van elk onderzoeksprogramma. De beoordeling van de commissie bestond uit een korte tekst met de belangrijkste bevindingen en een score op een Likert-type 5-puntsschaal op vier verschillende dimensies: kwaliteit (van de publicaties, gegeven de nationale en internationale prestaties in het desbetreffende vakgebied), productiviteit (aantal publicaties, gegeven de kwaliteit en de omvang van de groep), relevantie (voor de wetenschap en de maatschappij) en langetermijnlevensvatbaarheid (van de groep, gegeven de nationale en internationale wetenschappelijke competitie).

De evaluatie van 2001

De eveneens door de VSNU georganiseerde tweede evaluatieronde wijkt op twee belangrijke onderdelen van de eerste af. In de eerste plaats wordt nu met twee commissies gewerkt: één voor economie (waarbij inbegrepen econometrie en bedrijfseconomie) en één voor bedrijfskunde. De economiecommissie evalueerde zestig onderzoeksprogramma's van acht universiteiten en de bedrijfskunde-commissie negentien programma's van zes universiteiten. Een tweede opvallende verschil is het minder gedetailleerde niveau waarop de onderzoeksprestaties worden gerapporteerd. Naast de drie soorten dissertaties is er een categorie publicaties in internationale tijdschriften (onderverdeeld in gerefereerd en niet-gerefereerd), andere academische publicaties (boeken, hoofdstukken in boeken en bijdragen aan congresbundels) en Nederlandstalige publicaties (ook hier weer onderverdeeld in gerefereerd en niet-gerefereerd). Ten slotte onderscheidde men een categorie 'professionale publicaties en wetenschappelijke rapporten' waarin toegepast onderzoek en onderzoek in opdracht kon worden verantwoord. Een overzicht van de input- en outputvariabelen die in de 1995- en 2001-evaluaties zijn gebruikt staan weergegeven in tabel 1.

Tabel 1 Input en Output categorieën zoals geregistreerd in de twee visitatieronden

Input-categorieën 1995 en 2001	Inputs	Verklaring
Eerste geldstroom personeel	Aio's	Assistenten in opleiding (promovendi)
	Overige staf	Vast en tijdelijk onderzoekspersoneel
Tweede geldstroom personeel	Aio's	Promovendi gefinancierd door NWO
	Overige staf	Overig tijdelijk onderzoekspersoneel
Derde geldstroom personeel	Aio's	Promovendi
	Overige staf	Overig vast en tijdelijk onderzoekspersoneel

Output-categories 1995	Outputs
Dissertaties	Dissertaties voorbereid in eigen instelling (I), of extern (E)/voorbereid in eigen instelling (I), of geheel extern (E)
Wetenschappelijke publicaties	Tijdschriftartikelen (nationaal en internationaal) in zes categorieën: A-F (A hoogste kwaliteit, F laagste kwaliteit)
	Boeken (Nederlandse en internationale) in vijf categorieën: A-E
	Boekbijdragen (nationaal en internationaal) in vijf categorieën: A-E
Annotaties	Bibliografisch vastgelegde resultaten van wetenschappelijk werk
Professionele publicaties	Boeken, boekgedeelten, tijdschriftartikelen of technische publicaties gericht op verspreiding van kennis buiten de academische gemeenschap
Output-categorieën 2001	Outputs
Dissertaties	Dissertaties voorbereid in eigen instelling (I), of extern (E)/voorbereid in eigen instelling (I), of geheel extern (E)
Internationale wetenschappelijke publicaties	Artikelen in internationale tijdschriften
	Internationale boeken
	Bijdragen in internationale boeken
	Bijdrage in Internationale congresbundels
Nederlandse wetenschappelijke publicaties	Artikelen in Nederlandse tijdschriften
	Nederlandse boeken
	Bijdragen in Nederlandse boeken
Professionele publicaties	Onderzoeksrapporten

De evaluatiesystematiek van de commissies in 2001 lijkt op die van 1995 en bestaat eveneens uit vijf puntsscores op de dimensies kwaliteit, productiviteit, relevantie en langetermijnlevensvatbaarheid. Men maakt echter één groot verschil: de betekenis van relevantie wordt nu beperkt tot relevantie voor beleid of voor het beroep en sluit relevantie voor de wetenschap uit.

5. Relatie tussen peer review en de bibliometrische methode

De scores op de vier beoordelingsdimensies vertonen een significante samenhang (Pearson correlatie coëfficiënten tussen 0,628 en 0,796, $p < 0,01$). Common factor-analyse (zonder rotatie) levert één dimensie op voor de twee evaluatierondes (zie tabel 2).

Deze gemeenschappelijke factor verklaart in 1995 bijna tachtig procent van de variantie, en in 2001 bijna zestig procent. Zoals uit tabel 2 blijkt, is dat vooral te wijten aan de andere betekenis die in 2001 aan de factor 'relevantie' is gehecht. Om met een zo betrouwbaar mogelijke kwaliteit-sindicator verder te werken, hebben we een samengestelde kwaliteitsscore geconstrueerd als een gemiddelde van de scores op de factoren kwaliteit, productiviteit en levensvatbaarheid (hetgeen leidt tot betrouwbare dimensies, getuige de Cronbach alphas van 0,90, 0,82 en 0,85 voor respectievelijk de evaluatie 1995, economie 2001 en bedrijfskunde 2001).

Tabel 2 Common Factor Analyse van Commissie-oordelen

1995 Evaluatie		2001 Evaluaties	
		Economie	Bedrijfskunde
Common factor			
Eigenwaarde	3.161	2.350	2.374
Verklaarde variantie	79%	58%	59%
Evaluatiecriteria	Factorlading	Factorlading	Factorlading
Kwaliteit	0.921	0.838	0.899
Productiviteit	0.889	0.825	0.832
Relevantie	0.853	0.418	0.282
Levensvatbaarheid	0.892	0.891	0.892

We gebruiken de samengestelde kwaliteitsscore om de samenhang te onderzoeken met de bibliometrische gegevens. Daarvoor gebruiken we ordinale regressie (zie tabel 3). Het grotere aantal resultaatcategorieën in de eerste evaluatie leidt niet tot een betere voorspelling van de evaluatie-uitkomst van de commissie. In beide evaluaties blijkt het aantal internationale publicaties de beste voorspeller. Daarbij valt op dat boeken in de E-categorie (niet-gepubliceerde onderzoeksrapporten in 1995) en professionele publicaties in 2001 negatief samenhangen met het commissie-oordeel. Professionele publicaties blijken het oordeel van de evaluatiecommissies negatief te beïnvloeden.

Tabel 3 Ordinale Regressie van bibliometrische data op de samenvattende kwaliteitsoordelen van de reviewcommissies (afhankelijke variabele)

1995 Evaluatie		2001 Evaluatie	
Bibliometrische variabelen	Schatters (Wald) ¹⁾	Bibliometrische variabelen	Schatters (Wald) ²⁾
Dissertaties	0.072 (0.830)	Dissertaties	0.037 (0.679)
Artikelen A3)	0.296 (17.779)***	Internationale tijdschriftartikelen	0.027 (19.266)***
Artikelen B4)	0.130 (18.636)***	Internationale boeken	-0.024 (1.187)
Artikelen C	0.017 (1.988)	Internationale boekbijdragen	-0.005 (0.361)
Artikelen D	0.009 (0.657)	Internationale congresbundels	0.007 (2.428)*
Artikelen E	0.000 (0.001)	Nederlandse tijdschriftartikelen	0.006 (2.100)*
Artikelen F	0.020 (0.439)	Nederlandse boeken	-0.029 (0.605)
Boeken A	0.085 (0.132)	Nederlandse boekbijdragen	-0.003 (0.019)
Boeken B	-0.043 (0.110)	Professionele publicaties	-0.011 (8.148)***
Boeken C	0.064 (0.729)		
Boeken D	-0.065 (2.359)		
Boeken E	-0.019 (6.137)**		
Chi-Square	86.744 (p<0.000)	Chi-Square	58.713 (p<0.000)
Nagelkerke	0.665	Nagelkerke	0.587
n ⁵⁾	81	n ⁵⁾	68

***: $p < 0.01$; **: $p < 0.05$; *: $p < 0.10$.

- 1): Ordinale regressie met Logit link functie (gelijkmatig verdeelde categorieën in de afhankelijke variabele). De coëfficiënten van de onafhankelijke variabelen zijn herschaald op de waarde tussen 1 en 5 van de afhankelijke variabele (het samenvattend commissie-oordeel).
- 2): Ordinale regression met de Probit link functie (normaal verdeelde latente variabele)
- 3) Artikelen A is het aantal artikelen in internationale top-tijdschriften. VSNU hanteert in de eerste ronde een eigen lijst van A-tijdschriften.
- 4): Artikelen B is het aantal artikelen in internationale excellente tijdschriften (deze categorie is één kwaliteitsniveau lager dan de toptijdschriften) volgens de lijst samengesteld door de VSNU.
- 5): Het aantal in de analyse opgenomen programma's is kleiner dan het aantal beoordeelde programma's door ontbrekende data in enkele categorieën (bijvoorbeeld programma's die aangaven te zullen worden opgegeven kregen geen score voor levensvatbaarheid).

6. Factoren die invloed hebben op kwaliteit en productiviteit

Als kwaliteitsindicator gebruiken we de samengestelde kwaliteitsmaatstaf die reeds in de vorige paragraaf is toegevoegd. Om de productiviteit van onderzoek te meten, berekenen we de relatieve efficiëntie van elk onderzoeksprogramma in vergelijking met de overige programma's met behulp van Data Envelopment Analyse (DEA) (Cooper, Seiford, & Tone, 2000; García-Valderrama & Groot, 2002). Als outputvariabelen nemen we de variabelen die volgens tabel 4 significant samenhangen met kwaliteit. Voor de evaluatie van 1995 zijn dat de aantallen A- en B-publicaties, en voor de evaluatie in 2001 de aantallen internationale

tijdschriftpublicaties, internationale congresbijdragen en Nederlandse tijdschriftartikelen. De aantallen fte's promovendi en onderzoekers zijn gebruikt als inputvariabelen. De aldus verkregen DEA-efficiency scores correleren het beste met de commissiebeoordelingen op de dimensie productiviteit, hetgeen een aanwijzing is voor convergente validiteit van de DEA-scores. De relatie tussen kwaliteit en de programmamakenmerken wordt geanalyseerd met behulp van OLS-regressie. Dit is voor de analyse van de productiviteit niet mogelijk door de scheve en afgebroken verdeling van de DEA-scores. We gebruiken derhalve truncated Tobit regressie (met als 0 en 1 afkappingen van de DEA-scores). De resultaten staan weergegeven in tabel 4.

Tabel 4 Samenhang tussen kenmerken van onderzoeksprogramma's, onderzoekskwaliteit (t-waarden) en productiviteit (z-waarden)

1995 Evaluatie			2001 Evaluaties	
	OLS regressie Kwaliteit	Tobit regressie ¹⁾ Productiviteit	OLS regressie Kwaliteit	Tobit regressie ¹⁾ Productiviteit
Onafhankelijke variabelen	b	B	b	B
Constante	---	0,162 (1,969)**	---	0,361 (0,001)***
<i>Personeel</i>				
Totale formatie (fte)	0,500 (5,052)***	-0,000 -0,002	0,403 (3,468)***	(-0,202) (-1,989)**
% aio's	0,186 (1,698)*	0,224 (1,607)	-0,012 (-0,105)	0,285 (1,715)*
<i>Financieringsbronnen</i>				
% 2e geldstroom	-0,120 (-1,239)	-0,008 (-0,044)	0,225 (1,864)*	0,475 (2,216)**
% 3e geldstroom	0,011 (0,099)	0,039 (0,247)	-0,199 (-1,671)*	-0,036 (-0,251)
<i>Wetenschapsgebied</i>				
Economie ²⁾	-0,311 (-2,558)**	-0,188 (-3,396)***	-0,086 (-0,502)	-0,073 (-0,991)
Bedrijfskunde ²⁾	-0,313 (-2,444)**	-0,239 (-3,439)***	-0,034 (-0,195)	-0,025 (-0,348)
	F=7,380*** Adj. R ² =0,324 n=81	R ² = 0,321 R ² -adj. = 0,256 n=82	F=3,075*** Adj. R ² =0,146 n=74	R ² = 0,185 R ² -adj. = 0,092 n=69

***: $p < 0,01$; **: $p < 0,05$; *: $p < 0,10$.

1): Tobit regression truncated sample (0,1), quadratic hill climbing procedure.

2): Dummies voor wetenschapsgebied, uitgangspositie is econometrie en operations research.

Opmerking: het gebruik van log(totale formatie) leidt tot vergelijkbare resultaten.

Academische discipline

In de eerste evaluatie (1995) blijkt discipline samen te hangen met onderzoekskwaliteit en -productiviteit. Meer analytische disciplines, zoals econometrie en operations research, zijn productiever en leveren hogere onderzoekskwaliteit dan onderzoeksgroepen in economie en bedrijfskunde. De replicatie in 2001 bevestigt dit beeld echter niet: de factor discipline is insignificant geworden. De samenhang is dus hoogstens tijdelijk geweest. De evaluatie in 1995 was de eerste in zijn soort en deze leidde tot een kritische beoordeling van het onderzoek in vooral de bedrijfs-economie en bedrijfskunde. Het lijkt erop dat de onderzoeksprogramma's in deze disciplines hun prestaties na de eerste evaluatie op de voor de commissie relevante criteria aanzienlijk hebben verbeterd.

Omvang

De omvang van de onderzoeksgroep blijkt een belangrijke factor te zijn: in beide evaluaties hangt omvang positief samen met kwaliteit. Om te zien of dit verband ook niet-lineair is, hebben we een kwadratische term toegevoegd. Deze blijkt in beide evaluaties significant en negatief te zijn, hetgeen wijst op afnemende schaalopbrengsten ($t = -3,014$; $p < 0,004$ voor de evaluatie in 1995 en $t = -2,658$, $p < 0,010$ voor de evaluatie in 2001). Analyses met behulp van General Linear Models hebben geen optimale omvang aangetoond: het gaat hier om monotoon afnemende schaalopbrengsten: de toevoeging van één fte aan grotere groepen leidt tot een lagere marginale kwaliteitswinst dan aan kleinere groepen. Deze analyse doet overigens geen uitspraak over causaliteit: grote groepen zijn wellicht beter in staat om onderzoekers een omgeving te bieden die hen optimaal in staat stelt kwaliteit te leveren. Maar ook omgekeerd kan gelden dat onderzoeksgroepen die hogere kwaliteit leveren ook beter in staat zijn onderzoeksfondsen te verwerven.

Tot onze verrassing vinden we tevens een negatief verband tussen omvang van de onderzoeksgroep en productiviteit in de evaluatieronde van 2001 (ook in de eerste ronde vinden we een negatief verband, maar deze is niet significant). Grotere onderzoeksgroepen brengen over het algemeen betere kwaliteit voort, maar doen dit bij een lagere productiviteit.

Samenstelling van de staf en van financieringsbronnen

Beide factoren hebben geen stabiele en overtuigende invloed als we beide evaluatieronden tegelijk bekijken. Het aandeel promovendi hangt in de eerste evaluatie samen met kwaliteit en in de tweede evaluatie met productiviteit. De herkomst van financiering blijkt in de eerste evaluatie

weinig effect te sorteren, maar dit is anders in de tweede evaluatie. NWO financiering wordt steeds omvangrijker en zou in de tweede evaluatie wel degelijk meer gewicht in de schaal hebben kunnen leggen. Opvallend is dat een groter aandeel contractresearch op gespannen voet blijkt te staan met de kwaliteit van wetenschappelijke output van het programma. Dit komt overeen met resultaten van eerder onderzoek naar de invloed van contractfinanciering op de wetenschappelijke kwaliteit van onderzoek (Cherchye & vanden Abeele, 2005).

7. Conclusies

In dit artikel zijn twee vragen centraal gesteld: in welke mate leiden bibliometrische gegevens tot dezelfde oordelen over de kwaliteit van wetenschappelijk onderzoek dan peer review, en waarin onderscheiden succesvolle onderzoeksprogramma's zich van de minder succesvolle. We hebben voor het beantwoorden van deze vragen gebruik kunnen maken van bibliometrische gegevens en van de oordelen van reviewcommissies in twee evaluatieronden. Dit geeft ons een unieke kans op replicatie. Bibliometrische data blijken redelijk goede voorspellers van peer review-resultaten, alhoewel in elke ronde andere gegevens zijn gebruikt. De meest bruikbare voorspellers zijn publicaties in internationale gerefereerde toptijdschriften (de zogenaamde A- en B-publicaties), bijdragen in congresbundels en artikelen in gerefereerde Nederlandse tijdschriften. Omvang van de onderzoeksgroep is het enige kenmerk dat overtuigend en permanent onderscheid maakt tussen de prestaties van onderzoeksprogramma's. Omvang blijkt positief samen te hangen met kwaliteit en negatief met productiviteit. De mogelijkheid van replicatie laat ook zien dat andere kenmerken slechts tijdelijk voor verschillen zorgen. Daarbij is vooral opvallend dat de discipline in de tweede ronde geen duidelijk verschil meer maakt in kwaliteit en productiviteit van onderzoeksgroepen. Het is duidelijk dat vooral de bedrijfseconomische en bedrijfskundige onderzoeksprogramma's sterk zijn verbeterd tussen de eerste en tweede evaluatieronde. De komende evaluatie zal duidelijk moeten maken of deze trend zich in de recente jaren heeft doorgezet. ■

Prof. dr. T.L.C.M. Groot is hoogleraar Management Accounting bij de Faculteit Economie en Bedrijfskunde van de Vrije Universiteit te Amsterdam.

Literatuur

- Ballas, A., Theoharakis, V. (2003). Exploring diversity in accounting through faculty journal perceptions. *Contemporary Accounting Research*, 20(4, Winter), 619-644.
- Baron, J. N., Hannan, M. T. (1993). The impact of economics on contemporary sociology. *Journal of Economic Literature*, 32(3), 1111.
- Beyer, J. M. (1978). Editorial policies and practices among leading journals in four scientific fields. *Sociological Quarterly*, 19, 68-88.
- Brinn, A., Jones, M. J., Pendlebury, M. (1996). UK Accountants' Perceptions of Research Journal Quality. *Accounting and Business Research*(Summer), 265-278.
- Cherchye, L., vanden Abeele, P. (2005). On research efficiency. A micro-analysis of Dutch university research in Economics and Business Management. *Research Policy*, 34, 495-516.
- Cooper, W. W., Seiford, L. M., Tone, K. (2000). *Data Envelopment Analysis*. Norwell, Massachusetts: Kluwer Academic Publishers.
- Dwyer, P. D. (1994). Gender differences in the scholarly activities of accounting academics: an empirical investigation. *Issues in Accounting Education*, 9(2, Fall), 231-246.
- Van Fleet, D. D., McWilliams, A., Siegel, D. S. (2000). A theoretical and empirical analysis of journal rankings: The case of formal lists. *Journal of Management*, 26(5), 839-861.
- García-Valderrama, T., Groot, T. L. C. M. (2002). Controlling the Efficiency of University Research in the Netherlands. In S. A. Thore (Ed.), *Technology Commercialization: DEA and Related Analytical Methods for Evaluating the Use and Implementation of Technical Innovation* (pp. 147-182). Bosten/Dordrecht/London: Kluwer Academic Publishers.
- Groot, T. (1999). Budgetary reforms in the non-profit sector: a comparative analysis of experiences in Health Care and Higher Education in the Netherlands. *Financial Accountability and Management*, 15(3 en 4), 355-378.
- Hasselback, J. R., Reinstein, A., Schwan, E. S. (2000). *Benchmarks for evaluating the re-search productivity of accounting faculty*. Tallahassee, Florida: College of Business, Florida State University.
- Johnes, G. (1988). Determinants of Research Output in Economics Departments in British Universities. *Research Policy*, 17, 171-178.
- Konrad, A. M., Pfeffer, J. (1990). Do you get what you deserve? Factors affecting the relationship between productivity and pay. *Administrative Science Quarterly*, 35, 258-285.
- Kuhn, T. S. (1970). *The structure of scientific revolution* (2 ed.). Chicago: University of Chicago Press.
- Lodahl, J. B., Gordon, G. (1973). Differences between physical and social sciences in university graduate departments. *Research in Higher Education*, 1, 191-213.
- Loft, A., Jorissen, A., Walton, P. (2002). From newsletter to academic journal: Creating the *European Accounting Review*. *European Accounting Review*, 11(1), 43-75.
- Nederhof, A. J., van Raan, A. F. J. (1987). Peer Review and Bibliometric Indicators of Scientific Performance: A comparison of Cum Laude Doctorates with Ordinary Doctorates in Physics. *Scientometrics*, 11, 330-350.
- Nederhof, A. J., van Raan, A. F. J. (1993). A bibliometric analysis of six economics research groups: A comparison with peer review. *Research Policy*, 22, 353-368.
- Newman, J. M., Cooper, E. (1993). Determinants of academic recognition: The case of the Journal of Applied Psychology. *Journal of Applied Psychology*, 78(3), 518-526.
- Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review*, 18(4), 599-620.
- Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. *Research Policy*, 27, 95-107.
- Sterken, E., Leeflang, P. S. H. (2000). Een expertlijst van tijdschriften in de bedrijfswetenschappen. *Maandblad voor Accountancy en Bedrijfseconomie*, 74(4, april), 73-85.
- VSNU (1994). *Quality Assessment of Research, Protocol 1994* (No. PU/130113). Utrecht: Vereniging van Samenwerkende Nederlandse Universiteiten.
- VSNU (2002). *Assessment of Research Quality in Economics* (No. PU/130347). Utrecht: Association of the Universities in The Netherlands (VSNU).
- Westerheijden, D. F. (1996). *Een solide basis voor beslissingen, Effecten van onderzoeks-evaluaties door de VSNU* (Onderzoeksrapport voor de VSNU). Utrecht: Vereniging voor Samenwerkende Nederlandse Universiteiten.
- Whittington, G. (1993). The 1992 Research Assessment Exercise. *The British Accounting Review*, 25(4, December), 383-395.
- Whittington, G. (1997). The 1996 Research Assessment Exercise. *The British Accounting Review*, 29, 181-197.