Preliminary Draft not to be circulated (February 2017)

Examining the Returns to Investment in Science: A Case Study

Jacques Mairesse[1],Michele Pezzoni[2], Jacques Mairesse[1], Paula Stephan[3], Julia Lane[4]

[1] CREST-ENSAE, Paris, France; UNU-MERIT, Maastricht University, Netherlands; and NBER, Cambridge, Massachusetts, United States of America

[2] GREDEG, Nice University, Nice, France; CRIOS, Bocconi University, Milan, Italy; BRICK, Collegio Carlo Alberto, Torino, Italy

[3] Andrew Young School, Georgia State University, Atlanta, Georgia, and NBER, Cambridge Massachusetts, United States of America

[4] Wagner School, New York University, New York, New York, United States of America; BETA, University of Strasbourg, Strasbourg, France; University of Melbourne, Melbourne, Australia

Introduction

Program officers, asked to document the impact of their research funding, often draw a direct line between the grants they have funded and research output. Funding agencies do the same, aggregating grants together at the PI level. Yet there are multiple funders of scientific research and many recipients of scientific grants. Although this landscape reflects the complexity of modern knowledge production in which scientific research often requires multiple sources of funds, it also raises obvious questions for the management and evaluation of science funding. What is the link between an individual grant and subsequent research output? What are the effects of other sources of funding on the research output related to any given grant? These two questions must be addressed if program officers are to answer the fundamental question of the impact of a grant on the quantity and quality of research. Relating output to aggregate levels of funding must be addressed in order to answer the questions of the "optimal" quantity and size of grants. In addition to these questions, are the questions of whether research output related to a grant depends upon the size of the grant relative to other grants as well as the number of grants received by the PI.

This paper provides an approach to answering these questions. In doing so, three challenges must be addressed. Establishing causality is one, because grant funding and research output are endogenously determined. Researchers who receive grants are likely to differ in terms of creativity, ability and persistence from those who receive no or few grants. There is also the related endogeneity issue that at the PI level outcomes may be correlated overtime. One reason for such correlation is the process that Robert Merton called the Matthew Effect. Establishing attribution is another challenge because researchers may or may not directly acknowledge funders or inflate attributions, acknowledging multiple grants. The third challenge is the technical issue of relating the timing of the disbursement of grant funds with the timing of subsequent publications. Previous work has not had access to the data necessary to address the challenges, using only data from one funding agency, or using data that aggregate grants at the investigator level (*1*, *2*).

We are able to estimate the link between research funding and research output at the grant level using rich transaction level data at a small highly selective university for the period 2000-2010. After addressing the issues of endogeneity, attribution, and timing, we find that each 10% increase in the dollar size of a grant increases the quantity of publications by 4.3%. We similarly find that each 10% increase in the size of a grant increases the quality of publications by 2.7%. We find a negative relationship between the flow of funds from other grants and output: other things equal, an increase of 10% in funds from other grants leads to a decrease in publications associated with the focal grant of 1.9% and a decrease in quality of 2.0%.

Relationship between Funds and Productivity

We expect a relationship between the flow of funds from the focal grant and output given that grant funds permit the PI to devote resources to research, such as the PI's time and the time of research assistants and postdoctoral fellows. Funds also provide access to materials and equipment that, in the absence of the grant, would not be available. We also expect output associated with a focal grant to be related to funds received from other grants. These funds have potentially both a positive and a negative effect on focal grant output. The positive effect is that they provide resources which can be shared across research projects. The negative effect comes from the productivity expectation that accompanies the receipt of other funds which takes time away from the focal grant. An additional reason an increase in funds from other grants may lead to diminished output associated with the focal grant relates to what Jon Lorsch refers to as "bandwidth." Bigger labs, associated with more funds, mean that PIs have less time to devote to individual members of the lab. Moreover, more funds come with administrative tasks such as the preparation of proposals and the submission of progress reports. This administrative burden is well documented. A 2007 report of the Faculty Standing Committee of the Federal Demonstration Partnership (FDP) found that faculty working on federally supported research spent 42% of their research time on pre-and post-grant award administration activities—not on research(*3*) . While some of these bandwidth issues relate to the amount of funding, some of the administrative costs also relate to the number of grants.

Literature Review

The size and scale of the research enterprise is well documented. In the United States, the National Institutes of Health (NIH) awards over 9,000 research project grants (RPGs) annually to over 25,000 US researchers working in health-related research. The US National Science Foundation (NSF) makes about 11,000 grants. In Europe, the European Union awards about 7,000 grants a year to 42,000 researchers working in a wide array of areas, while national and regional agencies fund many more.

The impact of these research investments is less well understood(*4*, *5*), although science agencies have striven to directly tie research funding to research output. Different authors have attempted to address the attribution of publications to grants the timing of the disbursement of grant funds to the timing of publications; none have addressed the endogeneity issue.

In one of the most cited attempts to link research funding and grant output, Berg tied the amount of direct funding NIGMS investigators received in fiscal 2006 to the number of articles published during the period 2007-2010 and found a correlation coefficient of .14 between the two(*6*). He plotted the average number of publications, and their impact factor, of grant recipients to bins of funding and found that that researcher productivity began to diminish as grant size exceeded $600,000 to $750,000. He addressed the methodological issue of attribution by assuming that all articles published by the investigator during the period can be attributed to NIH funding, despite the fact that many investigators have funding from more than one agency. He addresses the methodological issue of timing by assuming that all articles published during the period 2007-2010 relate to funding received in 2006 although some of the research undoubtedly relates to funding received in an earlier or later period.   The Berg posting generated considerable discussion about the appropriateness of the output measure and the potential for conflating funding levels with lab characteristics.

In subsequent work, Berg found that researchers with one NIH grant produced a median of 11.5 articles; those with 2 produced a median of 23.5; the median for those with 3 was higher but the marginal impact was considerably lower than 11.5; as was the marginal impact of going from the 4th to 5th to 6th grants(*7*).

Lorsch, Berg's successor as director of NIGMS, built on this research, and a much earlier piece by Bruce Alberts(*8*), which outlined inefficiencies that can arise as labs become larger, to argue that inefficiencies are created when research funds are heavily concentrated among researchers rather than distributed more widely to the research community. The argument, simply stated, is that efficiencies arise by distributing funds more evenly (*11*). The conclusion is based on two different empirical analyses. One compares outputs attributed to NIH American Recovery and Reinvestment Act grants (ARRA) with outputs attributed to regular R01-funded NIH grants and found that "despite shorter durations and lower budgets, ARRA R01 grants had comparable citation outcomes per $million spent to that of contemporaneously funded payline R01 grants."(*9*) The other examined the publications of university researchers funded by the Natural Science and Engineering Research Council (NSERC) of Canada with those who received funds from a second Canadian funding council grant, and found the latter to be no more productive in terms of article counts than those who received only funds from NSERC(*10*).

In recent work Lorsch and coauthors examined the relationship between research funding and output for NIGMS investigators supported on research project grants and center grants[1]. Attribution is determined by citations to funding source in publications. Timing is addressed by relating funding in 2010 to publications in the four-year period 2011-2015(*12*). They find that publications increase proportionately with funding up to $300,000 but thereafter do not. They interpret this to mean that constant returns to scale are not present after $300,000. Their use of the terminology "constant returns to scale" is inconsistent with the meaning of the term in economics, where it relates to what happens to output when all inputs change proportionately. The inconsistency arises because, in the NIGMS case, only funding is changing; other inputs, such as lab space and major equipment, remain constant[2]. The authors also relate the average number of citations per publication to bins of funding and find that the number of citations per publication associated with a thousand dollars of funding decreases as funding increases.

Approach

We examine the relationship between funding and research output at the grant level. As noted above, there are three distinct issues to be addressed: the endogenous relationship between receiving a grant and researcher productivity, the attribution of publications to grants, and the timing of the disbursement of grant funds relative to the timing of subsequent publications.

We address the challenge of endogeneity by adopting a two-pronged approach. First, because the data are sufficiently rich to provide repeated observations on each faculty with different grants over time, we are able to include PI fixed effects in our analysis. These fixed effects control for unmeasurable characteristics such as ability, creativity and persistence that are invariant over time and relate to output. Second, we create an instrumental variable for the amount of funding which addresses the "dynamic" issues of endogeneity referred to above. The instrument is estimated based on the flow of funds at the national level, the flow of funds for the first year for the PI, the number of grants the PI has, and the source of the PI's funds. Instruments are created in this way for both the flow of funds from the focal grant as well as the flow of funds from other grants at the PI level.

We develop a new approach to link publications to grants given the unreliability associated with using citations to attribute publications to grant (*13*). The data (described in detail below) enable us to directly measure whether a student or postdcotral research(s) is supported on a grant (g) which also supports a faculty member. We then trace, through the Web of Science, all joint publications (p) between that faculty member and student or postdoc (see figure 1) and use this to attribute publications to a given

[1] NIGMS investigators were defined to be those supported on R01 equivalent or P01 grants in 2010.

[2] It is worth noting that their use of the term "constant returns to scale" is inconsistent with the meaning of the terminology in economics, where it relates to what happens to output when all inputs change proportionately. The inconsistency arises because, in the NIGMS case, only funding is changing; other inputs, such as lab space and major equipment, remain constant

grant. To be a bit more specific, we attribute articles to a grant if grant funds were used in a specific period to support graduate students as research assistants and/or as postdoctoral fellows and their names appear as co-authors along with the name of the faculty member on the publication.

Articles published at any time might be the result of research activity within a research grant (g) over several years. Or, they might be the result of research activity funded by other grants. We address these issues of timing in a more disaggregated fashion than have previous researchers. We note that, even though funding is obligated in a given year, simple attribution of activity to the year of obligation is likely to be misleading, since the disbursement of funds and the associated work occurs over the period of the grant. Therefore, we compute the flow of funds associated with a grant in any year to be the total direct costs of the grant divided by the number of years of funding provided by the grant. We calculate the flow of funds from each focal grant in each time period, and the flow of funds from all other grants in each time period. Our general methodology is to attribute articles to the flow of funds from grant g in year t if they are published in year t+1 and t+2. The lag between the flow of funds and publications that we choose reflects the fact that a period of one to two years often occurs between the performance of research and its publication. For purposes of robustness we estimate our models using alternative lag structure and other methodologies for determining attribution. See supporting material.

The basic model we estimate relates publications to the flow of funds from the focal grant and the flow of funds from all other grants during the period: The equation is specified in log form in order to estimate the elasticities.[3]

$$\log(Productivity) = \beta_0 + \beta_{grant}\ \log(\widehat{focal\ grant}) + \beta_{spillover}\log(\widehat{spillover})\ +\ controls_2\,\beta_2$$

Data

The panel contains 3796 observations at the PI-grant-year level. The data are built from the administrative records at the university, which, among other things, provide information on how the flow of funds was expended over time for each grant. We use this information to attribute articles to grants. Using the method outlined above, we compute a mean number of articles per PI-grant-year of 1.49 with a maximum 24.48 and a minimum of 0.04. [4]

The 1544 grants on average have a length of slightly over 4 years and a total awarded amount of $440,000. Approximately 33% come from NSF, 26% from NIH, 7% from DOD, 4% from DOE, and 31% from other sources.

---

[3] The log linear specification assumes a Cobb-Douglas production where $Y_{g,i,t} = (X_{g,i,t})^{\alpha} * (Z_{g,i,t})^{\beta}$. Our choice is based on the fact that estimation of a more general CES production function suggests that the Cobb-Douglas funcational form is appropriate.

[4] We recognize that this approach ignores that knowledge produced in a lab is embodied in multiple forms rather than being selectively parsed between outputs.

Table 1: Summary statistics for the study sample PI-Grant-year

	mean	sd	min	max	N
N. of publications attributed	1.49	1.76	0.04	24.48	3796
Average IF	1.98	2.37	0.02	36.10	3796
N. of pub*AVG if	3.79	7.32	0.00	96.13	3796
x flow of focal grant [M$]	0.11	0.09	0.01	0.52	3796
z spillovers from other grants [M$]	0.38	0.33	0.00	2.01	3796
Dummy one grant	0.07	0.26	0.00	1.00	3796
Dummy two grants	0.13	0.34	0.00	1.00	3796
Dummy three grants	0.16	0.36	0.00	1.00	3796
Dummy four grants	0.16	0.37	0.00	1.00	3796
Dummy five grants	0.13	0.33	0.00	1.00	3796
Dummy six grants	0.11	0.31	0.00	1.00	3796
Dummy seven or more grants	0.25	0.43	0.00	1.00	3796
National R&D investment US [M$]	60200.56	7716.42	40690.00	67367.00	3796
PI's flow in the first year [M$]	0.28	0.22	0.01	1.16	3796
N. of grants	5.03	3.09	1.00	17.00	3796
NSF grant	0.33	0.47	0.00	1.00	3796
NIH grant	0.26	0.44	0.00	1.00	3796
DOE grant	0.04	0.19	0.00	1.00	3796
DOD grant	0.07	0.25	0.00	1.00	3796
Other grant	0.31	0.46	0.00	1.00	3796
Department of Biology	0.17	0.38	0.00	1.00	3796
Department of Chemistry	0.21	0.41	0.00	1.00	3796
Department of Engineering	0.28	0.45	0.00	1.00	3796
Department of Geology	0.21	0.41	0.00	1.00	3796
Department of Physics, Math, Astronomy	0.13	0.34	0.00	1.00	3796

Results

Table 2 presents the regression findings. The dependent variable in models 1-3 is the log of publications attributed to the focal grant. Equation 1 controls for four variables: flow of funds at grant level, flow of funds from other grants at the PI-year level, time dummies and fixed effects. Because the log of zero is undefined, we impute a value of $50,000 to the flow of funds from other grants for the 7% of the observations for which the focal grant is the only grant at time t. Equation 2 instruments the flow of funds from the focal grant. Equation 3 instruments both the focal funds and the flow of funds from other grants.[5] In this, the preferred specification, the elasticity is estimated to be .43 indicating that a 10% increase in funds in a year from the focal grant is associated with a 4.3% increase in articles attributed to it in years t+1 and t+2. From the point of view of a program officer, this suggests that increasing the size of a PI's grant comes with positive returns. The coefficient on the log of the flow of funds from other grants is negative, suggesting that other things being equal an increase in funds from other grants is associated with a 1.9% decline in articles attributed to the focal grant. The intuition underlying the negative effect is that despite the fact that other funds can be expended on resources that complement the work of the focal grant, they also carry expectations of work as well as stretch what Lorch refers to as the bandwidth of the PI.

Equation 4-6 follow a similar strategy for estimating the elasticities with regard to the IF of articles associated with the grant. Preferred equation 6 in which both flows are instrumented indicates that a 10% increase in the flow of funds is accompanied by a 2.7% increase in the quality of articles. A 10% increase in the flow of funds from other grants comes with a 2.0% decrease in the quality of articles associated with the focal grant. The smaller focal grant elasticity observed for quality compared to quantity suggests that some resources in the research process are fixed, such as the ability of the PI, the PI's administrative skills or the quality of researchers the PI can attract to the lab.

Table 2: Regression Results. The dependent variables are the flow of articles attributed to a grant (columns 1-2-3) and their average impact factors (columns 4-5-6)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV(x)	IV(xz)	OLS	IV(x)	IV(xz)
VARIABLES	log(PhD/PD Pubs)	log(PhD/PD Pubs)	log(PhD/PD Pubs)	log(avg IF)	log(avg IF)	log(avg IF)
x (log flow at grant level)	0.33***	0.38***	0.43***	0.028	0.17*	0.27***
z (log spillovers at grant level)	-0.15***	-0.14***	-0.19***	-0.13***	-0.11***	-0.20***
time dummies	yes	yes	yes	yes	yes	yes
PI Fixed effects	yes	yes	yes	yes	yes	yes
Constant	-0.60	-0.21	-0.16	-0.30	1.63***	1.72***
Observations	3,796	3,796	3,796	3,796	3,796	3,796
R-squared	0.324	0.323	0.319	0.506	0.497	0.475

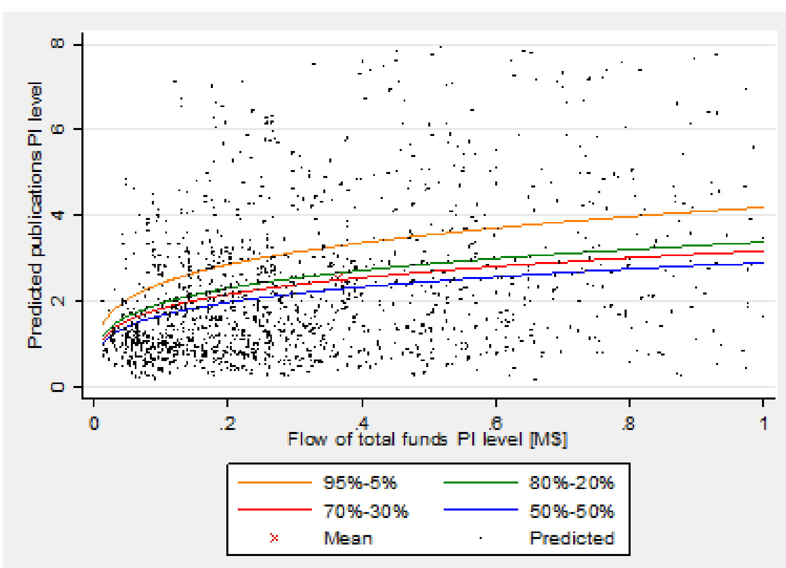
[5] The University has six divisions: Physics, Mat & Astronomy; Engineering and Applied Sciences; Chemistry and Chemical Engineering; Biology and Biological Engineering; Geological and Planetary Sciences; and Humanities and Social Sciences. We exclude the last division in our analysis and instead focus on the natural sciences and engineering.

Up to this point the analysis has focused on what happens to output when the flow of funds from a focal grant increases holding constant funds from other grants, or when the flow of funds from all other grants increases, holding constant funds from the focal grant. A related question is what happens to output when the flow of total funds increases. We see this as the policy question of importance to decision makers interested in the "big picture," such as directors of agencies and, more generally, directors of national offices such as OSTP. We interpret the coefficient on the focal grant to be of more interest to a program officer.

A straightforward answer to the question cannot directly be obtained by adding the coefficient on the flow of focal grant funds to the coefficient on the flow of other grant funds. This is because the functional form we have chosen means that the elasticity of output with regard to the total flow of funds depends upon the *share* that the focal grant represents in the PI's funding portfolio. To see this, requires a rewriting of equation (1) and manipulation of the terms so that the elasticity of publications with respect to total funds is equal to the sum of the coefficients on both flow variables minus the coefficient on the flow of other funds multiplied by the logit share of the focal grant. This means that when the two flows are equal, the elasticity is equal to the sum of the two coefficients. When the focal grant's share is larger than 50% the elasticity is the sum of these two coefficients plus a positive adjustment; if the focal grant's share is less than 50% the elasticity is the sum plus a negative adjustment.

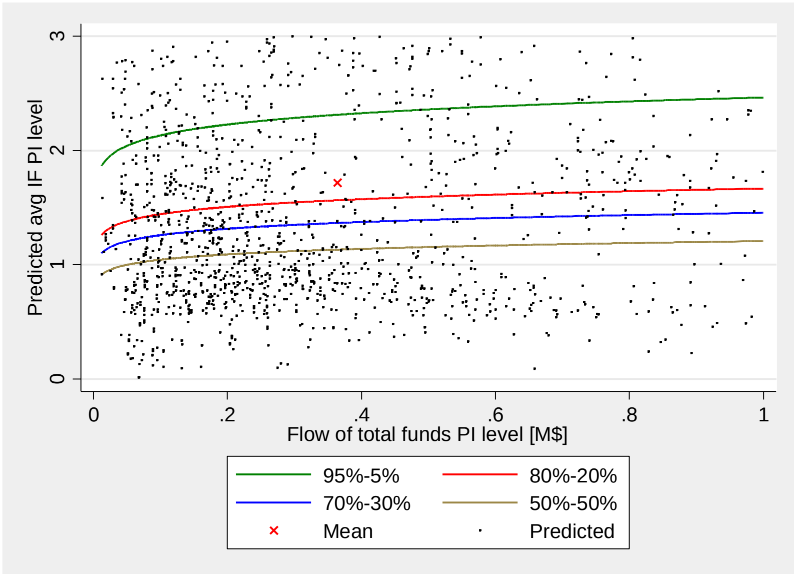
We use the insights provided by equation 3 to explore the relationship between publications and the flow of total funds for four shares of the focal grant. Figure 1 shows the relationship for quantity of publications; Figure 2 for quality of publications. Figure 1 shows that for any share, output increases as the flow increases and that for any total flow, output increases as the share of the focal grant increases.

Figure 1. Total flow at PI level VS. Count of publications BY share of the focal grant over total flow of funds at PI level. It is based on the estimation of column 3, Table 2



The related question, with regard to the Impact Factor, is explored in Figure 2. The findings are somewhat similar to those with regard to quantity. However, in this instance, output, as measured by the IF is particularly sensitive to the share of the focal grant, suggesting that there is a distinct advantage, holding funds constant, to allocating the majority of funds to one grant.

Figure 2: Total flow at PI level VS. average IF BY share of the focal grant over total flow of funds at PI level. It is based on the estimation of column 6, Table 2



Next we use the results to simulate the relationship between output and the flow of funds, varying the number of grants, assuming that all grants are of equal size (Figure 3). The algebra underlying the figures is shown in the supplementary material

Figure 3: Total flow at PI level VS. Count of publications BY number of grants of equal size. It is based on the estimation of column 3, Table 2

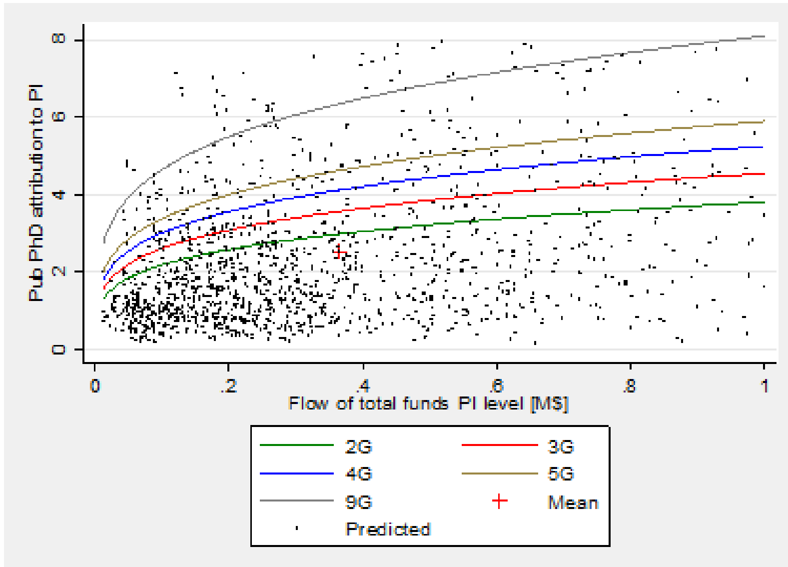
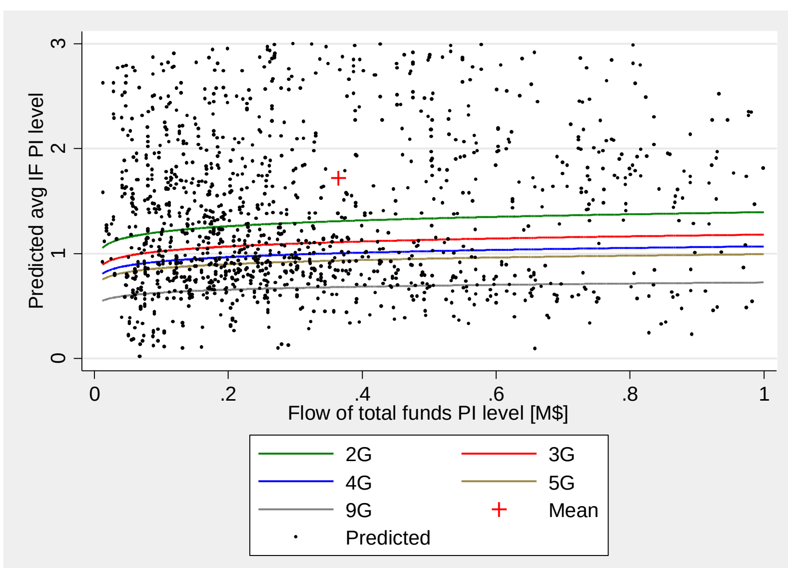


Figure 3 shows that output increases as the number of grants increase, holding flow of total funds constant. The finding is consistent with the expectation effect discussed above, suggesting that PIs have a strong incentive to produce at least one paper per grant. The same, however, cannot be said with regard to quality, as can be seen in Figure 4 which shows that holding funding constant, the quality of articles decreases as the number of grants increases. This suggests that there is a quantity quality tradeoff.

Figure 4: Total flow at PI level VS. average IF BY number of grants of equal size. It is based on the estimation of column 6, Table 2



Discussion

Using a novel method to attribute publications to grants we estimate the elasticity of publications to the flow of funds at the grant level. In order to control for endogeneity, we include fixed effects and instrument the flow of funds as well as the flow of funds from other grants. We refer to the latter as "spillovers". We find a significant and positive relationship between publications and flow of focal funds, suggesting that other things being equal an increase of 10 percent in a focal grant leads to a 4.3% percent increase in the total number of publications coming from that grant. Output declines as the PI gets more funds from other grants, consistent with Lorsch's bandwidth argument as well as what we refer to as the "expectation" effect. We find similar, although slightly muted results with regard to quality of publications.

When we aggregate funds across grants, we find a positive relationship between the flow of total funds and output as well as quality of output. However, the relationship is larger, other things being equal, the larger is the share of total funding devoted to the focal grant. The finding is congruent with the MIRA pilot initiative announced by NIGMS in July 2014 that supports an investigator's research through a single grant rather than through separate projects.(*14*) It is also congruent with the National Cancer Institute's newly established Outstanding Investigator Awards that will replace a PI's project-based grants with 7 years of funding of up to $600,000 a year in direct costs(*15*). Combining the focal grant coefficient with the coefficient for spillovers suggests that a 10% increase in the total flow of funds produces a 2.5% increase in the number of articles (0.50 -0.25). Our research also provides insight into whether, other things being equal, it is more efficient to have two grants of equal size or the same amount of total funding dispersed in four grants of equal size. For quantity, the answer is that more grants are associated with more output. For quality, the exact opposite is the case.

The positive and significant elasticizes that we estimate lend strong support to the contribution that funding makes to productivity—even the productivity of highly successful researchers at one of the US's top research institutions. Clearly funding facilitates productivity. Moreover, given the conservative methodology that we adopt for attributing publications to grants, which undoubtedly misses research facilitated by funding which did not directly support graduate students or postdocs, that these are "low ball" estimates.

It is our hope that our approach to estimating the relationship between research inputs and research outputs will be adopted and improved upon by researchers using the UMETRICS data that will soon become available at the IRIS institute at the University of Michigan. There is a clear need to know whether our findings are unique to a small institution with exceedingly high standards or apply to larger institutions with a bit more heterogeneity.

Bibliography

1. J. R. Lorsch, Maximizing the return on taxpayers' investments in fundamental biomedical research. *Mol. Biol. Cell*. **26**, 1578–1582 (2015).

2. J. Berg, Estimated publication outputs as a function of number of R01 grants per PI. *Datahound Blog* (2015), (available at http://datahound.scientopia).

3. R. Decker, L. Wimsall, A. Trice, J. Konstan, "A profile of Federal-Grant Administrative Burden among Federal Demonstration Partnership Faculty" (Washington DC, 2007).

4. C. MacIlwain, Science economics: What science is really worth. *Nature*. **465**, 682–684 (2010).

5. J. Lane, Let's make science metrics more scientific. *Nature*. **464**, 488–489 (2010).

6. J. Berg, Measuring the Scientific Output and Impact of NIGMS Grants (2010), (available at https://loop.nigms.nih.gov/2010/09/measuring-the-scientific-output-and-impact-of-nigms-grants/).

7. J. Berg, Another Look at Measuring the Scientific Output and Impact of NIGMS Grants (2010), (available at Another Look at Measuring the Scientific Output and Impact of NIGMS Grants).

8. B. M. Alberts, Limits to growth: in biology, small science is good science. *Cell*. **41**, 337–338 (1985).

9. N. S. Danthi, C. O. Wu, D. DiMichele, W. K. Hoots, M. S. Lauer, Citation Impact of NHLBI R01 Grants Funded Through the American Recovery and Reinvestment Act as Compared to R01 Grants Funded Through a Standard Payline. *Circ. Res.* **116**, 784–788 (2015).

10. J.-M. Fortin, D. J. Currie, Big Science vs. Little Science: How Scientific Impact Scales with Funding. *PLoS One*. **8**, e65263 (2013).

11. J. R. Lorsch, Lab Size: Is Bigger Better? *iBiology* (2015), (available at http://www.ibiology.org/ibiomagazine/jon-lorsch-lab-size-is-bigger-better.html).

12. T. Dorsey, J. Basson, J. R. Lorsch, Revisiting the Dependence of Scientific Productivity and Impact on Funding Level (2016), (available at https://loop.nigms.nih.gov/2016/07/revisiting-the-dependence-of-scientific-productivity-and-impact-on-funding-level/).

13. J. Rigby, Systematic grant and funding body acknowledgement data for publications: new dimensions and new controversies for research policy and evaluation. *Res. Eval.* **20**, 365–375 (2011).

14. NIGMS, Request for Information (RFI): Soliciting Comments on a Potential New Program for Research Funding by the National Institute of General Medical Sciences (2014), (available at http://grants.nih.gov/grants/guide/notice-files/NOT-GM-14-122.html).

15. J. Kaiser, NIH institute considers broad shift to "people"awards. *Science (80-.).* **345**, 366–367 (2014).

Supplementary Material (INCOMPLETE AND TO BE UPDATED): Tables S1 to S7

Table S1: Study sample

Initial sample	6822 (100%) - 283 PIs - 2167 Grants
Positive productivity grant-year pairs PhD/Postdoc attribution (Study sample)	3026 (44.4%) - 240 PI - 1544 Grants

Table S2: summary statistics at PI level (240 PIs)

	mean	sd	min	max	N
Number of publications PI-year (in year t)	8.45	7.85	0.00	68.00	1564
Number of publication PI-Year [PhD/Postdoc attribution]	3.62	4.10	0.05	44.71	1564
Flow of funds PI-year [M$]	0.26	0.24	0.01	1.78	1564
Department of Biology	0.17	0.37	0.00	1.00	240
Department of Chemistry	0.15	0.35	0.00	1.00	240
Department of Engineering	0.30	0.46	0.00	1.00	240
Department of Geology	0.18	0.38	0.00	1.00	240
Department of Physics, Math, Astronomy	0.20	0.40	0.00	1.00	240

Table S3: summary statistics at Grant-year level (1544 Grants)

	mean	sd	min	max	N
Number of publication Grant-Year [PhD/Postdoc attribution]	1.49	1.76	0.04	24.48	3796
Grant length	4.14	1.47	1.00	11.00	1544
Observation periods in the study sample	2.46	1.29	1.00	8.00	1544
Awarded amount	0.44	0.45	0.01	2.00	1544
NSF grant	0.23	0.42	0.00	1.00	1544
NIH grant	0.33	0.47	0.00	1.00	1544
DOE grant	0.34	0.47	0.00	1.00	1544
DOD grant	0.03	0.17	0.00	1.00	1544
Other grant	0.07	0.26	0.00	1.00	1544

Table SM4: Productivity versus number of grants

	N. of publications attributed	Average IF	N. of pub*AVG if	Obs.
Dummy one grant	2.13	2.34	5.62	275
Dummy two grants	1.64	2.14	4.18	496
Dummy three grants	1.51	2.26	4.40	593
Dummy four grants	1.39	2.23	4.06	610
Dummy five grants	1.22	2.20	3.38	484
Dummy six grants	1.36	1.90	3.52	404
Dummy seven or more grants	1.47	1.39	2.79	934

Table SM5: Regression Results: Dependent variable is flow of articles attributed to a grant

	(1)	(2)	(3)	(4)
	a	b	c	d
	log(PhD/PD Pubs)	log(PhD/PD Pubs)	log(PhD/PD Pubs)	log(PhD/PD Pubs)
	OLS	OLS	IV(x)	IV(x) and IV(z)
x (log flow at grant level)	**0.27*** **	**0.27*** **	**0.18*** **	**0.15*** **
z (log spillovers at grant level)	**-0.089*** **	**-0.12*** **	**-0.11*** **	**-0.088*** **
one grant (dummy)	0.48***	0.52***	0.52***	0.49***
BIOLOGY DIVISION		-0.10*	-0.058	-0.056
CHEMISTRY and CHEMICAL ENGINEERING		0.45***	0.46***	0.45***
ENGINEERING and APPLIED SCIENCE		-0.055	-0.048	-0.046
GEOLOGICAL and PLANETARY SCIENCES		-0.063	-0.094	-0.10*
PI Fixed effects	no	no	no	no
Constant	0.43***	0.34***	0.12	0.082
PI-Grant-year obs.	3,796	3,796	3,796	3,796
R^2	0.057	0.096	0.092	0.088

Table SM6: Regression Results: Dependent variable: average impact factors of the publications attributed with the PhD/Postdoc attribution method

	(1)	(2)	(3)	(4)
	a	b	c	d
	log(PhD/PD IF)	log(PhD/PD IF)	log(PhD/PD IF)	log(PhD/PD IF)
	OLS	OLS	IV(x)	IV(x) and IV(z)
x (log flow at grant level)	**0.21*** **	**0.11*** **	**0.26*** **	**0.30*** **
z (log spillovers at grant level)	**-0.016**	**-0.13*** **	**-0.13*** **	**-0.16*** **
one grant (dummy)	0.19***	0.33***	0.34***	0.37***
BIOLOGY DIVISION		0.77***	0.69***	0.69***
CHEMISTRY and CHEMICAL ENGINEERING		0.49***	0.47***	0.48***
ENGINEERING and APPLIED SCIENCE		-0.44***	-0.45***	-0.45***
GEOLOGICAL and PLANETARY SCIENCES		-0.12**	-0.064	-0.052
PI Fixed effects	no	no	no	no
Constant	0.73***	0.27***	0.65***	0.70***
PI-Grant-year obs.	3,796	3,796	3,796	3,796
R^2	0.034	0.233	0.217	0.209

Table SM7: First stage equation in the IV-2SLS estimation

	(1)	(2)
	x (log flow at grant level)	z (log spillovers at grant level)
	OLS	OLS
log(Flow of funds at national level)	**-0.062**	**0.27*** **
log(Flow in first year)	**-0.35**	**0.0036**
Number of grants	**-0.022*** **	**0.14*** **
NIH	**-0.039**	**-0.052**
DOE (Dep. of Energy)	**0.37*** **	**-0.061**
DOD (Dep. of Defense)	**0.0058**	**0.093* **
Other grants	**-0.24*** **	**-0.021**
BIOLOGY DIVISION	-1.87**	4.13***
CHEMISTRY and CHEMICAL ENGINEERING	-0.91	0.81
ENGINEERING and APPLIED SCIENCE	-0.099	0.58
GEOLOGICAL and PLANETARY SCIENCES	-0.54	0.35
PI Fixed effects	yes	Yes
Constant	-2.12	-7.18***
PI-Grant-year obs.	3,796	3,796
R^2	0.412	0.579