Bundling Matters:
Research Performance and Composition of Grant Portfolio

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

Draft: Do Not Circulate Without Authors' Permission

January 2020

[1] CREST-ENSAE, Paris, France; UNU-MERIT, Maastricht University, Netherlands, and NBER, Cambridge, Massachusetts, United States of America.
[2] GREDEG, Nice University, Nice, France; Observatoire des Sciences et Techniques, HCERES, Paris, France; ICRIOS, Bocconi University, Milan, 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.

Abstract: Using a novel method to attribute principal investigator's publications to each research grant in his or her portfolio, we estimate the elasticity of the grant flow of funds to the grant publication outcome, in terms of quantity and quality of the publications that can be directly linked to the grant. We find a significant and positive relationship between the flow of funds of a grant and the quantity of the publications that can be directly linked to the grant: an increase of 10% in the flow of funds leads to 4.3% and 2.6% increases in the number and average impact factor of the publications respectively coming from that grant. Aggregating these results at the level of principal investigators having many different grants at the same time, we can conduct two counter-factual simulations to assess whether and to what extent the bundling of grants affects publication productivity. Conditional on having the same amount of overall funding, we find there is a significant tradeoff between quantity and quality of publications. Having two grants of very unequal size is more efficient both in terms of quantity and quality; however, having more grants of the same size is associated with more publications of lower quality.

1. Introduction

Productive scientists often are supported on multiple grants covering their research expenses. This raises three important policy questions: (1) What is the return to a specific grant (herein referred to as the "focal" grant) given the presence of other grants? (2) How does the presence of other grants affect the return to the focal grant? And (3) For a given amount of total funding, what is the most efficient bundle of grants? By way of example, with two grants is it better to have a large one and a small one than it is to have two for the same amount? With multiple grants of the same amount, is it better to have fewer than more grants? Questions 1 and 2 are of concern to program officers seeking to evaluate returns to specific grants, while question 3 is of concern to policy makers seeking to design efficient funding structures for supporting scientists and scientists themselves assessing their performance.

In this paper, we estimate the return to a focal grant, holding funds from other grants constant, and how funds from other grants affect the returns to the focal grant. Our outcome measure is the number of publications attributed to a grant and a quality dimension of the publications measured by the average impact factor of the journals where the attributed publications appear. Data analyzed for public funding were provided by a small highly selective university in the US for all faculty from 2000 to 2010. Data include the identity of all the university faculty who are principal investigators (PIs) on at least one public research grant. The available information includes both grant and PI biographic characteristics. To measure the publication outcome, we complement this information with the PI's publications collected from the Web of Science (WOS) bibliometric dataset (Institute for scientific information, Thomson Reuters).

Our research is unique in that we measure the actual flow of funds from the focal grant rather than solely the presence of a grant or the amount of total funding a researcher has, as is typical in this type of research. We also address the extent to which funds from other grants affect outcomes related to the focal grant and how the way in which grants are bundled affects output as well as issues of attribution and endogeneity. We find that the elasticities of quantity and quality of publications attributed to a focal grant with respect to its size (as measured by the yearly flow of funds) are respectively 0.43 and 0.27, holding constant the flow of funds from other grants, meaning that a 10% increase in focal grant funds is associated with a 4.3% increase in number of publications and a 2.7% increase in the average impact factor of publications. We find negative elasticities of -0.19 and -0.20 for the quantity and quality of publications attributed to the focal grant with respect to the size of other grants. We attribute the negative elasticity to the additional commitment and expectations that come with additional grants. Both detract from effort directed to the focal grant and outweigh the positive effects provided by resources from other grants which can be shared across projects and which can provide funds for "lumpy" purchases such as equipment and postdoctoral scholars. These negative effects relate in part to the fixed costs associated with writing and administrating a proposal and are likely to be larger, the smaller and more numerous are other grants.

We create simulations based on these estimates to address the issue of bundling. We find that overall returns at the PI level are sensitive to the composition of a PI's grant portfolio. Specifically, we show that increasing the concentration of funding in one grant for a PI with two grants enhances the PI's productivity both in terms of quantity and quality for a given amount of total funding while increasing the number of grants for a PI with grants of equal size impacts productivity positively in terms of quantity but comes at the expense of quality.

In Section 2 we discuss the methodology developed to attribute PI's publications to a grant. In Section 3 we explain how we assess the relationship between the grant flow of funds and its attributed outcome and describe the data which we use. In Section 4 we present the estimates we thus obtain at the grant level. In Section 5 we evaluate the implications of the grant level estimates at the PI level, by considering two complementary simulations showing the effect of grant bundling on scientific productivity. Section 5 concludes.

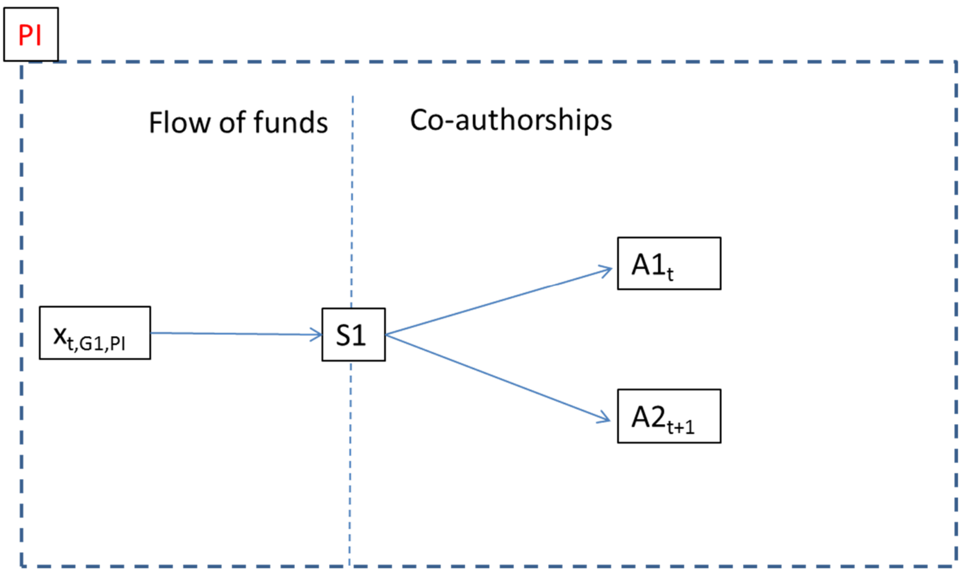
2. Attribution of the PI's publications to grants

We address the issue of the attribution of publications to different grants by developing a new approach, referred to as *Ph.D./PD attribution*. With this approach, we attribute a publication to a grant if the publication satisfies two conditions. First, the publication must be a co-authored research work between the PI and one of his or her PhD or Postdoc, second, the PI must have used his or her grant funds to support the PhD or Postdoc. We consider articles that are potentially attributable to the grant flow of funds in year t only those articles published in the three-year window covering year t and the two following years (t+1) and (t+2). The lag between the flow of funds and published articles reflects the fact that a period of one to two years might occur between the research activity conducted jointly by the PI and the funded PhD or Postdoc and the publication of its results. The *PhD/PD attribution* methodology allows us also to share the credit of one article among different flow of funds avoiding article double counting. Specifically, when the source of funding of the PhD or Postdoc is not unique, an equal share of the article credit is attributed to each of the multiple flow of funds. To be more precise, we can consider the two following examples.

Example 1. Attribution of multiple publications to a unique grant

Figure 1 shows how two articles, A1 and A2, are attributed to the flow of funds $X_{t,G1}$ of the grant *G1* awarded to the principal investigator PI. PhD student *S1* receives a salary in year *t* paid by the PI using funds from grant *G1*. At the same time *S1* is a co-author of PI on the publications A1 and A2. According to the *PhD/PD attribution* method, the full credit of both articles is assigned to the flow of funds of the grant *G1* in year *t*, namely $X_{t,G1}$.

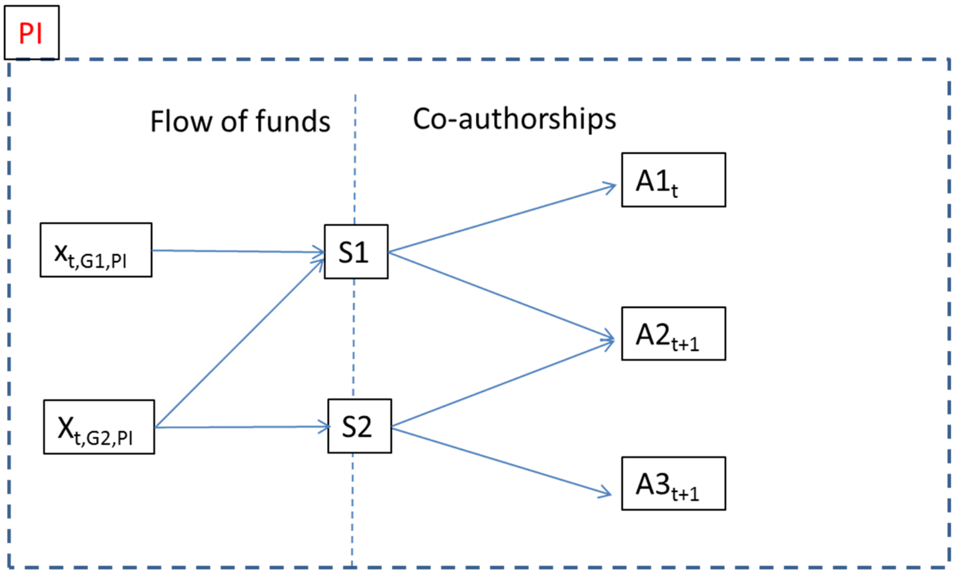
Figure 1: PhD/PD attribution method, example 1



Example 2. Attribution of multiple publications to multiple grants

Figure 2 shows how the *PhD/PD attribution* method assigns credit for three articles A1, A2, and A3 to two flows of funds ($X_{t,G1}$ and $X_{t,G2}$) from two different grants (*G1* and *G2*) both awarded to the same principal investigator PI. The salary of the PhD student *S1* is paid in year *t* by the PI on both grants *G1* and grant *G2*, while the salary of the PhD student *S2* is paid only on grant *G2*. The student *S1* is co-author of the PI on the articles A1 and A2, while *S2* is co-author of PI on the articles A2 and A3. The credit of the article A1 is shared equally by $X_{t,G1}$ and $X_{t,G2}$, namely half of the credit of article A1 is attributed to each of the two flows of funds since S1 is paid both by G1 and G2. One third of the article A2 is attributed to $X_{t,G2}$ through S2 and two thirds of A2 is attributed to $X_{t,G2}$ through S1 since G2 pays the salaries of both S1 and S2 (2/3 of the credit), while G1 pays the salary of S1 only (1/3 of the credit). The article A3 is fully attributed to the flow of funds $X_{t,G2}$ since G2 pays the salary of S2. In sum, the flow of funds $X_{t,G1}$ has a publication credit of 5/6 (1/2 (A1)+1/3 (A2)) articles, while the flow of funds $X_{t,G2}$ has a publication credit of 13/6 (1/2 (A1)+1/3 (A2)+1/3 (A2)+1 (A3)) articles. The sum of the credit attributed to both grants equals to 3 articles (5/6+13/6).

Figure 2: PhD/PD attribution method, example 2



According to the *PhD/PD attribution* method, we attribute 8,990 publications to 3,796 grant-year flows of funds included in our study sample. Table 1 shows the proportion of publications that are fully attributed to one flow of funds (unique grant and unique year), to multiple flows of funds of the same grant in different years (unique grant and multiple years), to multiple grants with simultaneous flows of funds in the same year *t* (multiple grants and a unique year), and to multiple grants and multiple flows of funds in different years (multiple grants and multiple years). Fifty-two percent of the credit of the publications included in our study sample is shared by multiple grants and cannot be attributed to the research expenses of a unique year.

Table 1: Proportion of publications according to their attribution

Publication credit attributed to:	Mean	Std. Dev.	Min	Max
A unique grant and a unique year	21%	0.41	0	1
A unique grant and multiple years	22%	0.44	0	1
Multiple grants and a unique year	5%	0.49	0	1
Multiple grants and multiple years	52%	0.50	0	1
	100%			

We have explored other attribution methods to attribute publications to grant-year flow of funds. For instance, we have used a purely chronological approach that attributes to each grant-year flow all the articles published by the PI in a following time window. Nonetheless, these alternative methods lead to noisy and unreliable attributions double counting the published articles and neglecting the issue of attribution when multiple grants are actively funding the PI research.[1] The *PhD/PD attribution* approach can be applied in our specific case due to the unprecedented level of detail of our data that enables us to know whether a Ph.D. or postdoctoral researcher is supported on a grant of a given PI. Considering the timing, we checked the robustness of the choice of considering as potentially attributable articles only those published in t, t+1, t+2. We find that our econometric estimations are robust also when considering other publication windows for determining attribution, i.e. t, t+1; t+1, t+2; and t, t+1, t+2, t+3.

[1] We have also tried to implement topic modelling to link publications to grants on the basis of a large corpus of grant abstracts. We found it to be impractical in our case, but certainly a method to develop in the future. We also considered using citations to funding sources reported in the publication acknowledgements, but the information was often either missing or overly vague (Rigby, 2011).

3. Methodology and data

To model the relationship between the grant flow of funds and its outcome, we consider Cobb–Douglas type regression as shown in Equation 1:

$$y_{g,i,t} = \alpha x_{g,i,t} + \beta z_{g,i,t}$$

Equation 1

where the subscript g stands for grant, i for PI, and t for year. The variable $y_{g,i,t}$ denotes the logarithm of $Y_{g,i,t}$, the number of articles attributed to the grant using *PhD/PD attribution* method (*Number of publications*) or, alternatively, the average five year impact factor of the journals where the attributed articles are published (*Average Impact Factor*). The variable $x_{g,i,t}$ and $z_{g,i,t}$ are respectively the logarithms of $X_{g,i,t}$ and $Z_{g,i,t}$, the flows of funds of the focal grant and the flow of funds of the other PI's grants active in year t. Finally, the coefficient α represents the elasticity of focal grant productivity with respect to the focal grant flow of funds, holding the flow of funds from other grants constant. The coefficient β is the elasticity of focal grant productivity with respect to the flow of funds from other grants, holding the flow of funds from the focal grant constant.[2]

We expect a positive relationship between the flow of funds of the focal grant and the grant publication outcome given that grant funds allow the PI to devote additional resources, such as the PI's time and the time of research assistants and postdoctoral fellows, to research associated with the focal grant. Focal grant funds can also provide access to materials and equipment that, in the absence of the grant, would not be available.

[2] The Cobb–Douglas type regression has an elasticity of substitution that equals to one, reflecting the fact that PIs in practice use funds from other grants to complement the funds from the focal grant. The choice of a linear regression would be unrealistic since it would assume an infinite elasticity of substitution. At the other extreme, a fixed proportion or Leontief production function would also be unrealistic by assuming a null elasticity of substitution. Furthermore, the choice of a Cobb-Douglas functional form appears as an excellent approximation. When we tested the more general Constant Elasticity of Substitution (CES) production function, we found an estimated elasticity of substitution very close to one for both the number of publications and the average impact factor equations.

We also expect the output associated with a focal grant to be related to funds received from other grants. Funds from other grants provide resources which can be shared across research projects. However, other grants also come with administrative tasks such as the preparation of proposals and the submission of progress reports that might negatively affect focal grant productivity. The administrative burden associated with grants is well documented in a report of the Faculty Standing Committee of the *Federal Demonstration Partnership* and in follow-up reports by the group in 2012 and 2019 that find that faculty working on federally supported research spend between 42% to 44% of their research time on pre-and post-grant award administration activities—not on research (Decker et al., 2007; Schneider, 2018).

A possible concern of estimating equation 1 regards the endogenous relationships between the grant flow of funds and the publication outcome of the PI. Researchers who receive grants are likely to differ in terms of ability, creativity and persistence from those who receive no or few grants. Assessing causality, is therefore, a major challenge, because grant funding and research output are endogenously determined. We can control for such unobserved differences in so far as they are largely permanent by taking advantage of the structure of our data to include PI fixed effects in estimating the relationship of publication to grant flow of funds. Introducing fixed effects can, however, aggravate estimation biases due to potential errors in variables and to other sources of endogeneity arising from the correlation of researchers' publication productivity over time, largely reflecting the process of cumulative advantage that Robert Merton called the Matthew Effect (Merton, 1968).

We address the endogeneity issue by instrumenting the flow of funds from both the focal grant and from other grants active in year t. The excluded instruments used in our analysis are the *growth rate of national funding* available for research projects, the *number of grants* the PI has, a dummy variable measuring whether the PI has funds from *more than 2 funding agencies*, and dummy variables for each funding agency. Appendix A reports a detailed description of each of the explanatory variables included in our IV first stage regression.

Descriptive statistics

Tables 2 and 3 present the descriptive statistics for the unbalanced PI-year level sample and for the unbalanced grant-year level sample, respectively. The average number of yearly publications by PIs receiving one or more grants is 3.62, varying between 0.05 and 44.71. The corresponding average impact factor is 2.37. The average *Total flow of funds* ($F_{i,t}$) is 0.36 million dollars per year. PIs receive 3.63 grants a year ($n_{i,t}$) on average from 1.49 funding agencies. In 37% of the PI-year observations, funds come from more than one agency. Table 3 shows that the average number of publications per year attributed to a grant equals 1.49, ranging from a minimum of 0.04 to a maximum of 24.48. The average impact factor is 1.98. The average flow of funds from a focal grant is 0.11 million dollars per year, while the average flow of funds from other grants is 0.38 million dollars per year.

Table 2: Descriptive statistics for unbalanced PI*year level sample (1564 observations, 240 PIs and 10 years)

Average per PI and year	mean	sd	min	max	N
Number of publications	3.62	4.10	0.05	44.71	1564
Average Impact Factor	2.37	2.47	0.18	31.02	1564
Total flow of funds ($F_{i,t}$) [M$]	0.36	0.30	0.01	2.03	1564
Number of grants ($n_{i,t}$)	3.63	2.43	1	17	1564
Number of funding agencies	1.49	0.74	1	5	1564
Proportion of grants from more than one funding agency	0.37	0.30	0	1	1564

Table 3: Descriptive statistics for unbalanced grant*year level sample (3786 observations for 1544 grants and 10 years)

Average per focal grant and year	mean	sd	min	Max	N
Number of publications	1.49	1.76	0.04	24.48	3796
Average Impact Factor	1.98	2.37	0.02	36.10	3796

4. Estimates at the grant level

Table 4 shows the estimation of the first stage equation of the instrumental variable approach. In Column 1, the dependent variable is the flow of funds of the focal grant (x), while in Column 2, the dependent variable is the flow of funds of the other grants (z). According to our expectations the *Growth rate of national funding* is positively correlated with focal grant flow of funds (x), while the variables *Number of grants* and *More than 2 funding agencies* are negatively correlated with focal grant flow of funds. The variable other grant flow of funds (z) is positively correlated with the variables *number of grants* and *More than 2 funding agencies.* We find that the focal grant flow (x) correlates with the funding agency awarding the grant.

Table 4: First stage equation in the IV-2SLS estimation

	(1) OLS x (log focal grant flow of funds)	(2) OLS z (log other grant flow of funds)
Growth rate of national funding	0.53***	-0.11
Number of grants	-0.025***	0.25***
More than 2 funding agencies	-0.068**	0.35***
NSF	0.25***	-0.026
NIH	0.21***	-0.093**
DOE (Dep. of Energy)	0.61***	-0.045
DOD (Dep. of Defense)	0.27***	0.010
Other funding agencies	ref.	ref.
PI Fixed effects	Yes	Yes
PI-Grant-year obs.	3,796	3,796
R^2	0.412	0.579

Table 5: OLS and IV estimates at grant level with PI fixed effects

VARIABLES	log(number of publications)			log(average impact factor)		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV(x)	IV(xz)	OLS	IV(x)	IV(xz)
x (log focal grant flow of funds)	0.33***	0.38***	0.43***	0.028	0.17*	0.27***
z (log other grants flow of funds)	-0.15***	-0.14***	-0.19***	-0.13***	-0.11***	-0.20***
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
PI Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
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
Sargan test (P value)	-	0.0002	0.0008	-	0.0000	0.0807

Table 5 shows the results of the estimated coefficients of Equation 1. The dependent variable of the regressions reported in Columns 1, 2 and 3 is the logarithm of the number of publications attributed to the focal grant, while the dependent variable of Columns 4, 5 and 6 is the average impact factor of the journals where these publications appeared. Columns 1 and 4 reports OLS estimates, Columns 2 and 4 use the predicted values of Table 4 Column 1 to instrument the focal flow of funds (x), while Columns 3 and 4 instrument both the focal flow of funds (x) and the flow of funds from other active grants (z) using the predicted values of Table 4 Column 1 and 2, respectively. We find that a 10% increase in the dollar flow of the focal grant increases the quantity and quality of publications attributed to the grant by respectively 4.3% and 2.7%, conditional on holding the size of the other grants constant. We also find a negative elasticity of the quantity and quality of the publications attributed to the focal grant with respect to other grants respectively of 1.9% and 2.0% for a 10% increase in the dollar flow of the other grants. This result is conditional on holding the flow of funds from the focal grant constant.

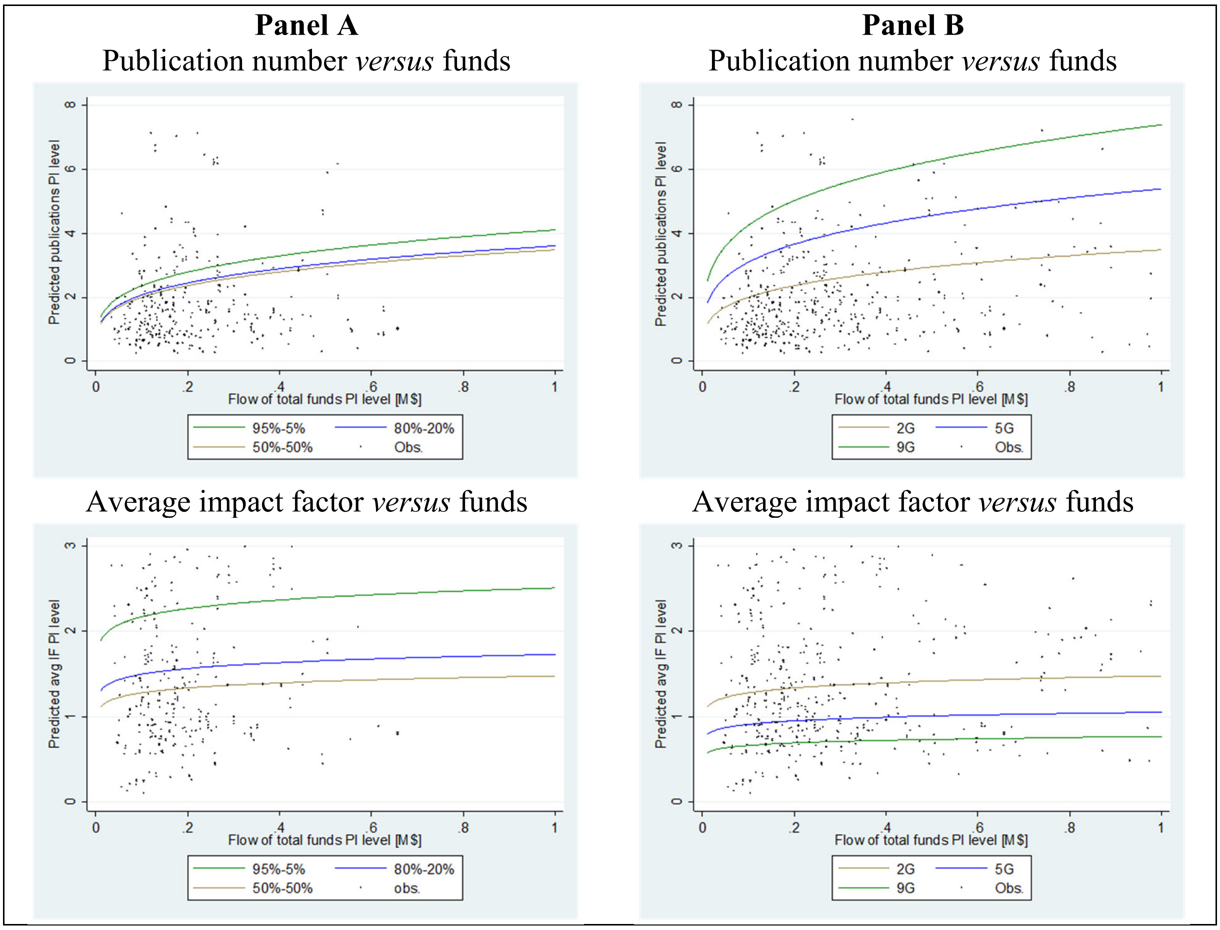
5. Evidence implied on bundling at the PI level

Relying on the results of our estimation at the grant-year level, we examine whether the quantity and quality of the overall PI's productivity relates to the way in which funding is received in terms of number and size of grants, i.e. the bundling effect. To do so, we start from the estimates of the elasticities α and β reported in Table 5, Columns 3 and 4, and we conduct two types of counterfactual simulations. The first simulation assumes that the PI has only two grants ($n_{i,t} = 2$) of different size and allows us to investigate the relationships between the predicted PI's publication productivity and the total flow of funding ($F_{i,t}$), varying the relative size of the two grants (s_1 and s_2). The second simulation assumes that the PI has multiple grants ($n_{i,t} > 1$) of equal size ($s=1/n_{i,t}$) and allows us to investigate the relationships between the predicted PI's productivity, the flow of individual funding, and the number of grants. Appendix B reports the details of the empirical strategy followed to use the results obtained at the grant-year level to simulate the productivity at the PI-year level by assuming different configurations of the PI's grant portfolio.

In terms of quantity, our priors are that due to administrative responsibilities associated with a grant, other things being equal, PI's will be more productive the more concentrated is their total funding. However, the total output will increase as the number of grants increases, due to the incentives to report to funders research results for each grant. Our priors with regard to quality are different. Although we expect quality to increase the more concentrated is funding, we expect it to decrease as the number of grants increases, reflecting the need to produce output related to each grant at the expense of quality.

Figure 3 presents the first set of simulations. Panel A assumes that the PI has only two grants with different size configurations: equal size (50%-50%), one is larger than the other (80%-20%), and one is practically equal to total funding (95%-5%). Figure 3, Panel A, shows that for any size of total funds, the number of publications (upper panel) and average impact factor (lower panel) increase as the share of one grant increases relatively to the other grant. The second set of simulations, Panel B, assumes that the PI has multiple grants of equal size. The number of grants consider ranges from 2 to 9. This simulation shows that, holding funding constant, more grants are associated with more output, but it comes at the expense of quality indicating a tradeoff between publication quality and quantity.

Figure 3: Simulated publication productivity and average impact factor as function of funding at PI level



6. Conclusion

Using a novel method to attribute publications to grants we estimate the elasticity of publications to the flow of funds at the grant level. We find a significant and positive relationship between the publications attributed to the focal grant and the corresponding flow of funds, implying that holding the flow of funds from other grants constant an increase of 10% in a focal grant leads to a 4.3% increase in the number of publications coming from the grant. We also find that a 10% increase in the flow of funds from other grants results in a 1.9% decline of the number of publications attributed to the focal grant holding the flow of funds from the focal grant constant. We find similar, although slightly muted results with regard to quality of publications as measured by the average impact factor of the journals in which articles are published.

We consider two sets of counterfactual simulations to investigate the extent to which the bundling of grants affects output. They show respectively that for the same amount of funding it is more efficient to have two grants of very unequal size for both quantity and quality of publications, while more grants of the same size are associated with more publications but of a lesser average impact factor. In short, in the case of multiple grants of equal size, a tradeoff exists between quantity and quality. These findings are 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 (NIGMS, 2014) and with its extension in 2017 to all NIGMS investigators, be they early stage, new or established.[3]

Methodologically, the PhD/PD attribution method presented here demonstrates that associating publications with grants can lead to inflation in the number of publications attributed to a grant if multiple grants are not taken into consideration. Specifically, our methodology shows that the majority of publications are associated with multiple grants over single or multiple years, not a single grant over a single or multiple year.

There are three main policy implications from our findings. First, program officers who want to increase the effectiveness of grants awarded by their funding agencies should consider the negative effects caused by grants from their own as well as funding agencies. Second, policy makers who design ways in which funding is distributed to PIs, and PIs performing research, should recognize

[3] See https://www.nigms.nih.gov/Research/mechanisms/MIRA/Pages/default.aspx.

that the bundling of grants can create a publication quality-quantity tradeoff. Our research methodology does not permit an investigation of other key policy issues related to the structure of grants, such as the duration of the grant and whether it is preferable to support projects or people.

It is our hope that our approach to estimate the relationship between research inputs and research outputs will be adopted and improved upon by researchers using the UMETRICS data that are becoming 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, selective, institution with exceedingly high standards or apply to larger institutions with more heterogeneity. Our prior is that the tradeoff we observe would be stronger if we were to study a more heterogeneous group of researchers rather than a highly selective group whose ability and expertise arguably attenuate the tradeoff.

References

Decker, R.S., Wimsatt, L., Trice, A.G., Konstan, J.A., 2007. A profile of federal-grant administrative burden among Federal Demonstration Partnership faculty. A Report of the Faculty Standing Committee of the Federal Demonstration Partnership.

Merton, R.K., 1968. The Matthew Effect in Science. Science 159, 56–63.

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

Rigby, J., 2011. Systematic Grant and Funding Body Acknowledgement Data for Publications: New Dimensions and New Controversies for Research Policy and Evaluation. Research Evaluation 20, 365–375.

Schneider, S., 2018. FDP Faculty Workload Survey: Plan & Preliminary Results 14.

Appendix A: Instrumental Variable estimation, excluded instruments

This appendix describes four excluded instruments used in the first step of the IV estimation: (1) *Growth rate of national funding*, (2) *Number of grants*, (3) a dummy variable *More than 2 funding agencies*, and (4) a set of dummies identifying the *funding agency* of the focal grant.

1. *Growth rate of national funding*

We expect growth rate of national funding to be positively correlated with the amount of the grants awarded to the PI in our study sample. To calculate the *Growth rate of national funding* variable we proceed in three steps. First, we consider five categories of flow of funds at national level as reported in the official statistics: NIH life sciences; other life sciences; physical sciences; engineering; mathematics and computer science.[4] We match at least one category to each department of the university where we conducted our study. The matching criterion is based on the relevance of the funding category at national level for the department. For instance, we expect funds in the category Engineering to be relevant for the PIs affiliated to the Department of engineering and to be not relevant for the other departments. Table A.1 shows the result of the matching. Second, we calculate the growth rate (positive of negative) of the funds for each department smoothed over three years, from t-2 to t. Figure A.1 shows the values of the growth rate for each department. Third, we calculate the variable *Growth rate of national funding* by assigning a growth rate value to each focal grant according to the department of affiliation of the awarded PI.

Table A.1: Matching between the department of affiliation of the PI and the classes of national funds that are considered relevant for the department.

University departments	Categories of flow of funds at national level
Department of Biology	NIH life sciences + other life sciences
Department of Chemistry	Physical sciences
Department of Engineering	Engineering
Department of Geology	Physical sciences
Department of Physics, Math, Astronomy	Physical sciences + math & computer science

[4] see https://www.aaas.org/page/historical-trends-federal-rd

Table A.2 shows the correlation between *Growth rate of national funding*, the focal grant flow of funds, and the other grant flow of funds. We expect also the *Growth rate of national funding* to be uncorrelated with the error term in the grant productivity equation.

Figure A.1: Growth of the funds at national level for each department

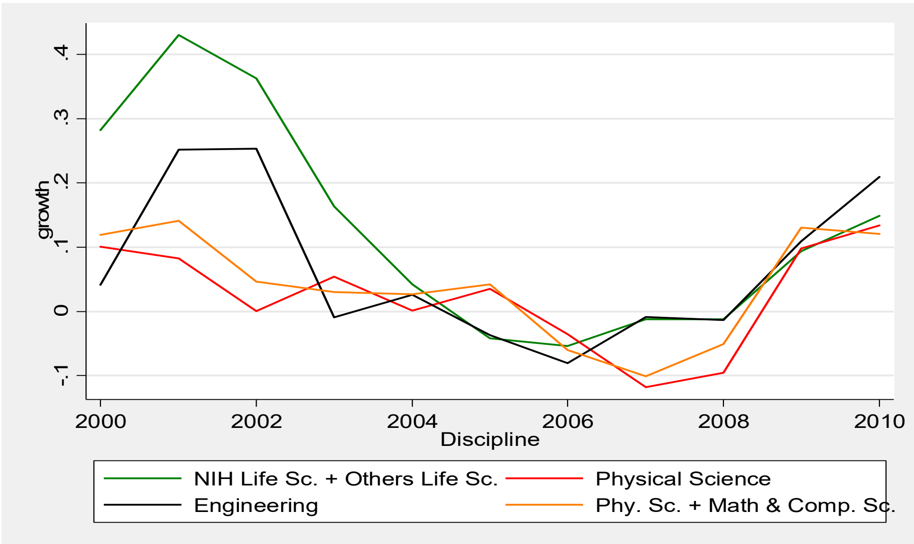


Table A.2: Correlation between growth and flow focal /other grants

	Growth rate of national funding	x	z
Growth rateof national funding	1		
x (Focal grant flow of funds)	0.14	1	
z (Other grant flow of funds)	0.04	0.12	1

2. *The number of grants*

The number of "active" grants in year *t* is expected to be positively correlated with the PI's other grant flow of funds in year *t*, i.e., the more the grants awarded the more the funds available. The number of active grants is a good instrument to correct for the measurement error problem affecting the flow of funds in year *t*. Table A.3 shows the correlation between the number of active grants in year *t*, the flow of funds of the focal grant, and the PI's other grant flow of funds.

Table A.3: Correlation between the number of grants in year t and the flow of funds of the focal grant and the PI's other grant flow of funds

	Number of grants	x	z
Number of grants	1		
x (Focal grant flow of funds)	-0.16	1	
z (Other grant flow of funds)	0.43	0.12	1

3. Dummy *More than 2 funding agencies*

The number of distinct funding agencies awarding the PI's other grant flow of funds, i.e., *z*, is a proxy for the ability of the PI to raise funds from different sources. Table A.4 shows the co-occurrence of different funding agencies in the PI's grant portfolio for our study sample. The values reported in the table represent the proportion of observations where, conditional on observing a focal grant awarded by a funding agency A, there is another grant in the PI's portfolio awarded by a funding agency B. For instance, when we consider a focal grant awarded by NIH, in 37% of the cases there is another grant awarded by NSF in the PI's grant portfolio. We define the variable *More than 2 funding agencies* as a dummy that equals one if there are two or more distinct agencies awarding the funds of the PI's other grants

Table A.4: Co-occurrence of the funding agencies in the PI's grant portfolio

		Other grant funding agency (B)					
		Only focal grant	NSF	NIH	DOE	DOD	OTHER
Focal grant Finding agency (A)	DOD	0.07	0.56	0.22	0.10	0.54	0.52
	DOE	0.08	0.69	0.26	0.30	0.19	0.64
	NIH	0.07	0.37	0.84	0.07	0.10	0.21
	NSF	0.11	0.67	0.20	0.10	0.18	0.51
	OTH	0.04	0.68	0.13	0.10	0.16	0.80

Table A.5 shows the correlation between the dummy *more than 2 funding agencies,* the flow of funds of the focal grant, and the PI's other grant flow of funds.

Table A.5: Correlation between the dummy more than two funding agencies, the flow of funds of the focal grant, and the PI's other grants' flow of funds

	More than 2 agencies	x	z
More than 2 agencies	1		
x (Focal grant flow of funds)	-0.08	1	
z (Other grant flow of funds)	0.26	0.11	1

4. *Dummy variables for each funding agency (NSF, NIH, DOE, DOD, others)*

Dummies identifying the funding agencies awarding the focal grants are expected to be correlated with the size of the focal grants (x). For instance, DOE and NIH tend to award grants of larger size than NSF and DOD. Table A.6 shows the average size of the grants included in our study sample by funding agency. The funding agency dummies are expected to be uncorrelated with the error term in the grant productivity equation.

Table A.6: Average grant size by funding agency

Funding agency	flow of funds of the focal grant [M$]
DOD	0.12
DOE	0.13
NIH	0.16
NSF	0.09
OTHERS	0.08

Appendix B: Using grant-year level estimations to investigate the bundling effect

Starting from the grant level estimates of the elasticities α and β reported in Table 5, we consider two types of counterfactual simulations.

The first simulation assumes that the PI has only two grants of different size and allows us to investigate the relationships between the predicted PI's publication productivity and the total flow of funding at the PI level ($F_{i,t}$), varying the relative size of the two grants. For instance, PI *i* in year *t* has available a total amount $F_{i,t}$ of funding for her research. $F_{i,t}$ is the sum of the flow of funds of two grants, 1 and 2. The grant 1 contributes to the PI's funding $F_{i,t}$ with the share s_1, while 2 contributes with the remaining share, i.e., ($s_2=1-s_1$). In the simulation we assume that the PI has two grants of equal size (50%-50%), one is larger than the other (80%-20%), and one is practically equal to total funding (95%-5%). The PI's publication productivity is calculated as in Equation B1, where α and β are the elasticities reported in Table 5, Column 3 and 6. When using the coefficients estimated in Column 3 we predict the PI's *Number of publications*, while when using the coefficients estimated in Column 4 we predict the *Average Impact factor* of the journals where the PI's publications appear.

$$Publication\,productivity_{i,t} = e^{y_{A,i,t}} + e^{y_{B,i,t}} = e^{\alpha\cdot s_1\cdot F_{i,t}} \cdot e^{\beta\cdot(1-s_1)\cdot F_{i,t}} + e^{\alpha\cdot(1-s_1)\cdot F} \cdot e^{\beta\cdot s_1\cdot F_{i,t}}$$

(Equation B1)

The second simulation assumes that the PI has multiple grants ($n_{i,t} > 1$) of equal size ($s=1/n_{i,t}$) and allows us to investigate the relationships between the predicted PI's productivity, the flow of individual funding, and the number of grants. In our simulation we assume that the PI has a varying the number of grants of equal size ranging from 2 to 9 grants ($n_{i,t}$). The PI's publication productivity is calculated as in Equation B2.

$$Publication\,productivity_{i,t} = n_{i,t} * e^{y_{.,i,t}} = n_{i,t} \cdot e^{\alpha\cdot(F_{i,t}/n_{i,t})} \cdot e^{\beta\cdot(n_{i,t}-1)\cdot(F_{i,t}/n_{i,t})}$$

(Equation B2)