

UC Santa Barbara

Econ 196 Honors Thesis

Title

Effect of Research Funding on the Number of Science & Engineering Doctorate Recipients: Evidence from the American Recovery and Reinvestment Act

Permalink

<https://escholarship.org/uc/item/2xd6w960>

Author

Chen, Marty

Publication Date

2019

Undergraduate

Effect of Research Funding on the Number of Science & Engineering Doctorate Recipients: Evidence from the American Recovery and Reinvestment Act

Marty Chen
Advisor: Youssef Benzarti

March 2019

Abstract

Federal support to universities and colleges in science and engineering is critically important because it affects the educational outcomes of future scientists. This paper aims to examine the effect of federal funding to U.S. universities on the corresponding science and engineering doctorate recipients. The main research question is: What's the impact of research funding on the number of science and engineering doctorate recipients, using evidence from the American Recovery and Reinvestment Act of 2009? Specifically, I utilize the difference-in-difference approach to compare, before and after ARRA, the number of doctorate recipients in fields most affected and not as affected by federal research obligations. Under two different settings of control and treatment groups, I find that fields which experienced large growth in funding during ARRA period did not show an increase in the number of recipients in the following years.

1. Introduction and Motivation

Looking at the history of science, a major shift occurred during the 20th century; the start of the World wars and the Cold war created the demands for performing extensive scientific research, and the pattern of university research gradually switched from basic research to applied research. Researchers respond more to social or government needs than to individual interests; and the topic of the research is often determined by the source of funds: government agencies, business companies, etc. University research continues this pattern into the 21st century, while scientists play an important role in innovation and fostering economic growth. Therefore, the science and engineering doctorate recipients are one of the key factors to study when attempting to understand the welfare of the whole economy.

The American Recovery and Reinvestment Act (ARRA) was an economic stimulus package enacted in 2009 that provided funds to many areas, including scientific research, to help the U.S. economy recover from the recession. Figure 1 shows R&D funding to U.S. universities from 2000 to 2016 by the top seven agencies¹ that covers over 97% of total

¹ Departments of Agriculture, Defense, Energy, Health and Human Services, and Homeland Security; National Aeronautics and Space Administration; and National Science Foundation.

funding. The post Great Recession years of 2009 and 2010 saw a large and sudden growth in research funding, which reached its peak in 2010 at 29 billion dollars.

However, after the ARRA period, federal support for university R&D has been gradually declining, and universities are increasingly relying on self-funding and funding from philanthropists. It was not until the latest data regarding the year of 2016 that showed an increase in federal science and engineering support; the preliminary data for 2017 also demonstrated a jump in the funding from 2016. The fact that investments in this area have been going up and down, to some extent, indicates that the effect of such investments is not clearly understood. If this effect was well-analyzed, the research funding over time would be more, in case the social return was positive and high, or less, in case the social return was low.

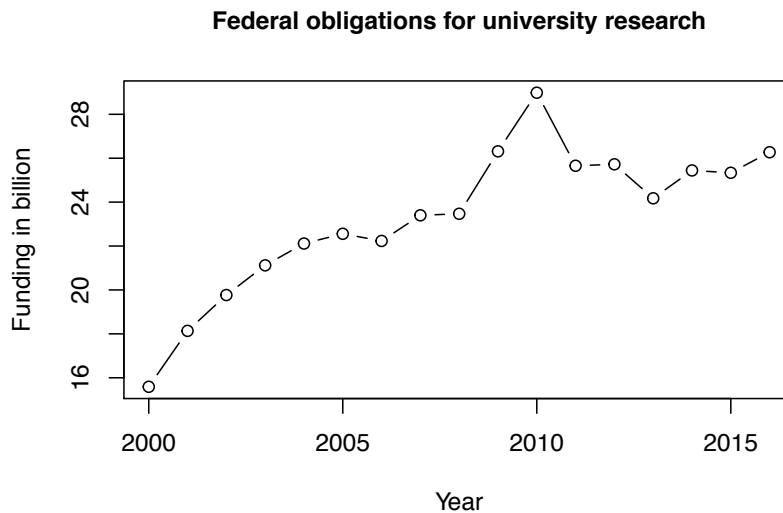


Figure 1: Federal obligations for research performed at universities and colleges for selected agencies

Previous studies² largely focused on the changing patterns of higher education R&D funding and short-term economic activities – employment, workforce decision, earnings, etc. – of PhD students who received federal funding. Quality of these students and their post-graduation decisions are examined by many; however, none has worked on the quantity of doctorate recipients – the magnitude of the effect of ARRA. Therefore, this paper aims to fill that gap and contribute to a comprehensive understanding of the impact brought by research funding to universities.

This paper uses a difference-in-difference method in regressing the effect of ARRA money on the number of science and engineering doctorate recipients. Data on research funding to universities and colleges will be matched to data on the number of recipients by subfields. Of those subfields that have data on both sides (for the rest of the paper I will refer to these subfields as “fields”), I separate them into control and treatment groups using two different criteria, each resulting from a distinct dataset. There are no hard evidence showing that The American Recovery and Reinvestment Act targeted a specific group of fields, rather it stated that “funding for new principal investigators and high risk, high return research were top

² See next section for details.

priorities.”³ Therefore, we apply an alternative method in determining the control and treatment groups: on one hand, fields that demonstrated a larger annual increase rate in research funding during the ARRA period than that of before ARRA period will be categorized as treatment group; on the other hand, fields that had smaller annual increase rate during ARRA period will be assigned to the control group.

The regression results show this effect not significant in both control-and-treatment settings, namely there is no significant difference in the patterns of doctorate recipients between fields in the control and treatment groups. The current data on research funding during ARRA years has a limited number of academic subfields; therefore, the amount of observations is relatively small in this paper and the coefficients involve large variation. However, the results of this paper still provide empirical evidence to policy makers who in the future may attempt to influence education outcomes through putting money into specific academic fields.

2. Literature Review

Past literatures documenting the effect of federal funding emphasized on post-graduation activities of students who received funding during ARRA period. *Federal Funding of Doctoral Recipients: Results from New Linked Survey and Transaction* in 2017 “provides new insights into how survey data can be combined with administrative records to examine the ways in which funding affects workforce decisions.”⁴ Both UMETRICS dataset, which is generated from university payroll and financial records, and the Survey of Earned Doctorates (SED), which is an important US survey dataset regarding the doctoral workforce, are studied and analyzed to obtain results that cannot be derived from any one of them. The UMETRICS project, which “contains record level information on wage payments made from federal grants to doctoral students and other university personnel,”⁵ is one of the main datasets that economists use when studying the effect of ARRA. Apart from these, the literature finds large differences in funding patterns across federal funding agencies in terms of the number and disciplinary training of doctoral students; also, federal funding is strongly related to individual characteristics and career pathways. Another paper, *Science Funding and Short-Term Economic Activity*⁶ in 2014 introduces the STAR METRICS project, on which the UMETRICS was built, and which documents not just short-term, but also longer-term, results of scientific activity in response to the ARRA in 2009. It has a clear mission to “provide policymakers with a better understanding of the process of research and provide the research community with a common

³ Bement L, Arden. *American Recovery and Reinvestment Act of 2009*. Retrieved from <https://www.nsf.gov/pubs/issuances/in131.jsp>.

⁴ Chang, Wan-Ying. Cheng, Wei. Lane, Julia. Weinberg, Bruce. *Federal Funding of Doctoral Recipients: Results from New Linked Survey and Transaction Data*. National Bureau of Economic Research, 2017.

⁵ Chang, Wan-Ying. Cheng, Wei. Lane, Julia. Weinberg, Bruce. *Federal Funding of Doctoral Recipients: Results from New Linked Survey and Transaction Data*. National Bureau of Economic Research, 2017.

⁶ Weinberg, Bruce. Owen-Smith, Jason. Rosen F, Rebecca. Schwarz, Lou. Allen McFadden, Barbara. Weiss E, Roy. Lane, Julia. *Science Funding and Short-Term Economic Activity*. *Science* 344 (6179), 41-43, 2014.

data infrastructure that connects research funding with research outcomes.”⁷ Along with UMETRICS, STAR METRICS is also an essential source of data on recipients of research funds.

One of the papers that utilize these major projects data is *Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. in 2015*. For eight universities, it combines the data from UMETRICS project and the U.S. Census Bureau and “covers 2010–2012 earnings and placement outcomes of people receiving doctorates in 2009–2011.”⁸ They match data on students who received ARRA funding and data on employment outcomes to examine the pattern of these recipients entering the workforce. The paper finds that almost 40% of supported doctorate recipients, entered industry and, when they did, they disproportionately got jobs at large and high-wage establishments in high-tech and professional service industries. It also observes geographic clustering in employment near the universities that trained and employed the researchers, as well as large differences across fields in placement outcomes.

My paper, similar to the previous paper that combines UMETRICS project and U.S. Census Bureau data, will match each of the two federal funding data- *Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions, and Higher Education Research and Development Survey (HERD)*- to data on science and engineering doctorate recipients in obtaining regression results with a difference-in-difference set up. Due to the limited accessibility of both UMETRICS and STAR METRICS projects, this paper does not extract data from these widely used datasets; instead, the National Center for Science and Engineering Statistics under the National Science Foundation, which will be introduced in the data section, provides us with rich data sufficient to the completion of this research. While all previous literatures discussed the educational or employment outcomes of students who received funding, none of them studied the magnitude of this effect in terms of quantity: how many students are affected by ARRA money being put into different academic fields? Or is there an effect of federal research funding on the number of students graduated in the following years? If the answer to the latter question is positive, it may provide the government with alternative methods when attracting younger workforce in designated fields.

3. Empirical Model

(1) Regression Model and Parameters

This paper utilizes a difference-in-difference approach in answering the research question. It best fits this context since we are able to measure the difference in the response of control and treatment groups to ARRA money in terms of the number of recipients. The Ordinary Least Squares(OLS) regression is

$$Y = \beta_0 + \beta_1 *ARRA + \beta_2 *ARRA_affect + \beta_3 *ARRA *ARRA_affect + \varepsilon,$$

⁷ Chang, Wan-Ying. Cheng, Wei. Lane, Julia. Weinberg, Bruce. *Federal Funding of Doctoral Recipients: Results from New Linked Survey and Transaction Data*. National Bureau of Economic Research, 2017.

⁸ Zolas, Nikolas. Goldschlag, Nathan. Jarmin, Ron. Stephan, Paula. Owen-Smith, Jason. Rosen F, Rebecca. Allen McFadden, Barbara. Weinberg, Bruce. Lane, Julia. *Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients*. Science 350 (6266), 1367-1371, 2015.

where the dependent variable Y is the number of science and engineering recipients of a specific academic field, $ARRA$ is a dummy variable which is equal to 1 if the observation is post-ARRA period and 0 if it is pre-ARRA, $ARRA_affect$ is another dummy variable which is equal to 1 if the selected field is heavily affected by ARRA money and 0 if it is not. The $ARRA_affect$ separates the observations into control ($ARRA_affect = 0$) and treatment groups ($ARRA_affect = 1$). Finally, there is an interaction variable $ARRA*ARRA_affect$, which is the product of $ARRA$ and $ARRA_affect$. As a result, the coefficient of the interaction term (β_3) would be the difference-in-difference estimation of the effect of ARRA.

(2) Control & Treatment group

The first decision in obtaining the regression parameters is the selection of control and treatment group. The mindset of this process is that: fields that are targeted by the ARRA money will be put into the treatment group since they are affected by the presence of ARRA, and fields that are not targeted will be in the control group to reflect a natural shift in the number of doctorate recipients. However, there are no hard evidence documenting any fields being targeted by the recovery act: the government papers merely state that the top priorities in allocating ARRA money would be high-risk, high-return projects. Therefore, this paper selects the control and treatment groups by looking at the annual increase rate of research funding received by a field during ARRA years. Specifically, if the annual increase rate of money received during ARRA period is more than the annual increase rate before ARRA, I claim this field as affected by recovery act funding and assign it to the treatment group; if the annual increase rate of money received during ARRA period is less than that of before ARRA, I claim this field as not affected by recovery act funding and assign it to the control group. The reasoning behind this is that if the annual increase rate of research funding during ARRA period, which I set to be $R1$, is greater than the annual increase rate before ARRA, which I set to be $R2$, then the presence of ARRA positively shifted the former trend of research funding to this specific academic field; and this is directly proportional to the stimulus package provided by the federal government.

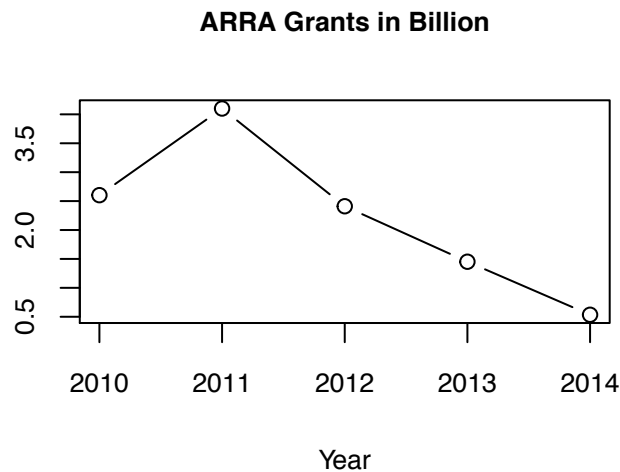


Figure 2: Federally financed higher education R&D expenditures funded by ARRA

Now, to determine the exact time frame of pre-ARRA and ARRA period, I look at the years that ARRA money was most in effect. Figure 2⁹ shows the higher education R&D expenditure funded by ARRA over the years. According to this data, ARRA funding was at its peak in 2011, when the R&D expenditures funded by ARRA was at \$4.1 billion. It is also observed that ARRA money was high during the first three years of accessible data, and gradually dropped to a low level starting from 2013, when the level of funding was below \$1.5 billion. Therefore, I define the “ARRA-period” as a three-year period from 2009 to 2012, and to match this time interval I define “pre-ARRA” as another three-year period from 2005-2008¹⁰. This allows me to capture the effect of ARRA during its peak years. To combine with the control and treatment selection mechanism discussed above and to sum up, I set the annual increase rate of research funding to a selected field from 2009 to 2012 as $R1$ and set the annual increase rate of this field from 2005-2008 as $R2$; by comparing the magnitude of $R1$ and $R2$, I assign this observation to the treatment group when $R1 > R2$ and to the control group when $R1 < R2$.

Federal financed higher education R&D expenditure

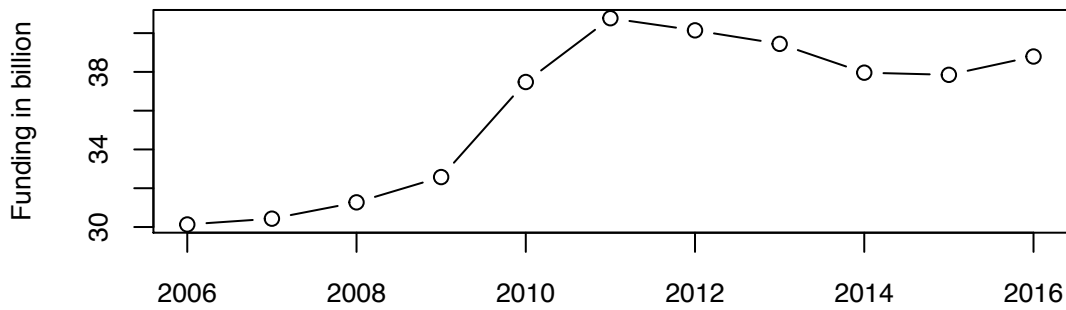


Figure 3: Setting 1 Funding Trend

Federal obligations for research performed at universities and colleges

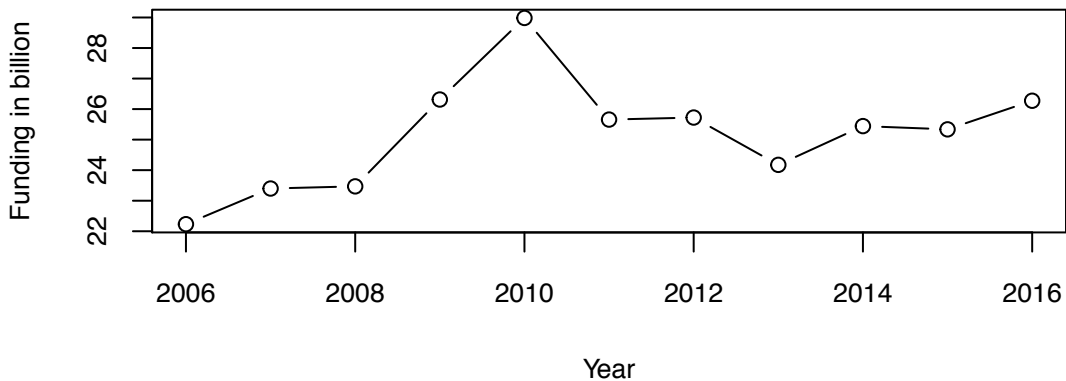


Figure 4: Setting 2 Funding Trend

⁹ Data obtained from the *Higher Education Research and Development Survey (HERD)*.

¹⁰ Figure 1 shows a quite smooth pattern of research funding during 2005-2008, meaning that the Great Recession does not seem to have a big impact on the amount of research funding. This allows me to select 2005-2008 as pre-ARRA period

After setting up the mechanism for determining control and treatment groups, I decide to regress the ARRA effect under two different settings, each resulting from a distinct set of datasets. The difference between these two settings is that setting 1 uses the *Higher Education Research and Development Survey (HERD)* dataset and setting 2 uses the *Survey of Federal Funds for Research and Development*. Figure 3 and 4 shows the research funding trend from 2006-2016 for setting 1 and setting 2. A noticeable difference is the year of 2011, when the federal financed higher education R&D expenditure (setting 1) increases but the federal obligations for research performed at universities and colleges (setting 2) decreased from 2010 by a large ratio. This seemingly contradictory discontinuity in data might be the result of a lag in measuring the level of funding. Therefore, I test the difference-in-difference estimator using two different settings of control and treatment groups to gain confidence in the regression results. If the regressions under these two different settings yield similar results, then we are confident to say our results are reflecting the actual effect.

In setting 1, I obtain 21 fields from matching up the HERD and recipients data; only 3 of them go into the control group, and the rest 18 are put into the treatment group¹¹. In setting 2, 20 academic fields have data on both the funding and recipients sides; this gives us 8 fields in the control group and 12 in the treatment group¹².

(3) The Timing of ARRA effect

Another main decision to make is the determination of the timing of ARRA effect, namely when will the *ARRA* dummy be equal to 1 and when will it be equal to 0. As discussed above, it should be 1 for observations on recipients from after ARRA and 0 for those from before ARRA. The question here is: how long will the ARRA money take into effect in terms of the number of science and engineering doctorate recipients? First, I determine the timing for pre-ARRA; to eliminate the effect ARRA years could possibly have on the number of recipients and to reduce variation that results from selecting only a single year, I set the average number of doctorate graduates from year 2007 to 2009 as the number of recipients pre-ARRA period. Then, I specify the timing for post-ARRA. Here I make the assumption that students on average graduate 3-5 years after they received research funding. Since the ARRA money was mostly distributed in 2009-2012, I set the average number of doctorate graduates from year 2014-2016 as the number of recipients post-ARRA period¹³. This assumption and definition are, to some extent, vague and might distort our desired effect. Therefore, I will perform a few robustness checks using different combinations of time frames to test the reliability of our regression results.

¹¹ Control group: Aeronautical Engineering, Oceanography, Political Sciences. Treatment group: Computer Sciences, Mathematics, Bioengineering and Biomedical Engineering, Chemical Engineering, Civil Engineering, Electrical Engineering, Mechanical Engineering, Metallurgy and Materials Engineering, Atmospheric Sciences, Agricultural Sciences, Astronomy, Chemistry, Physics, Psychology, Economics, Sociology, Education, Humanities.

¹² Control group: Mathematics, Chemical Engineering, Civil Engineering, Metallurgy and Materials Engineering, Agricultural Sciences, Psychology, Economics, Political Sciences. Treatment group: Computer Sciences, Aeronautical Engineering, Electrical Engineering, Mechanical Engineering, Atmospheric Sciences, Geological Sciences, Oceanography, Astronomy, Chemistry, Physics, Anthropology, Sociology.

¹³ This is obtained by adding up those 3-5 years to the ARRA period; also, the latest data available is for year 2016.

4. Data

(1) Data Description

This paper uses a number of survey data from the National Science Foundation (NSF). The National Center for Science and Engineering Statistics (NCSES), established within the NSF, is a provider of statistical data on the U.S. science and engineering enterprise, and is also the source of data for all the datasets listed below.

The first dataset is the *Higher Education Research and Development Survey (HERD)*. It is the primary source of information on R&D expenditures at U.S. colleges and universities. The survey collects information on R&D expenditures by field of research and source of funds and also gathers information on types of research, expenses, and headcounts of R&D personnel. Specifically, within the dataset, I will be using *Higher education R&D expenditures by source of fund and R&D field* to distinguish the fields that are heavily affected by federal funds, and thus ARRA. This can also be done by combining *Higher education R&D expenditures within standard form population* by R&D field and federally financed expenditures by R&D field. Moreover, the *Higher education R&D expenditures funded by the American Recovery and Reinvestment Act of 2009* is used to illustrate the trend of ARRA funds to university research over the years; this table is only available in HERD for the years between 2010 and 2014, for the fact that ARRA funds are mostly distributed during these years.

The second dataset is the *Survey of Federal Funds for Research and Development*. As an important source of information about federal funding for R&D in the United States, this survey is an annual census completed by the federal agencies that conduct R&D programs. Historically, the Federal funds data are collected annually for 3 government fiscal years at a time: the fiscal year just completed, the current fiscal year, and the next fiscal year. Actual data are collected for the year just completed and estimates are obtained for the current and next fiscal year. In this way, the trends of federal funds and comparison between adjacent years can be easily found and performed. Specifically, I will use the *Research obligations to university and college performers by selected agency and field of science and engineering, and by selected agency and detailed field*, which gives me more detailed data on federal funding to universities and the agencies that provide the money.

While the two datasets above show the patterns of science and engineering research funding to U.S. universities, the fourth dataset responds to the number of doctorate recipients in multiple settings. The *Survey of Earned Doctorates (SED)*, which began in 1957, is a survey administered to all the doctorate recipients in the United States and is sponsored by the NCSES and five other federal agencies¹⁴. "It collects information on the doctoral recipient's educational history, demographic characteristics, and post-graduation plans. The SED contains, inter alia, the following information: the doctorate's name, birth year, country of birth, race, sex, academic institution of the doctorate, sources of financial support during graduate school and sources and type of financial support for postdoctoral study or research."¹⁵ Specifically, I

¹⁴ National Institutes of Health, Department of Education, Department of Agriculture, National Endowment for the Humanities, and National Aeronautics and Space Administration.

¹⁵ Chang, Wan-Ying. Cheng, Wei. Lane, Julia. Weinberg, Bruce. *Federal Funding of Doctoral Recipients: Results from New Linked Survey and Transaction Data*. National Bureau of Economic Research, 2017.

will use *Doctorate Recipients, field and demographic characteristics, by subfield* to gather data on the number of doctorate recipients in specific fields, and match them with the fields selected in the previous datasets.

(2) Summary Statistics

A. Dependent Variable: The outcome variable in this difference-in-difference regression is the number of science and engineering doctorate recipients for each subfield. In the first set of control and treatment groups, where I match *Higher Education Research and Development Survey (HERD)* to the *Survey of Earned Doctorates (SED)*, there are 21 fields with data on both surveys, in which only three of them- Aeronautical Engineering, Oceanography, and Political Science- are in the control group. Since each field provides 2 observations, namely the number of recipients before and after ARRA, this gives us 6 observations in the control group and 36 observations in the treatment group.

Figure 5 shows the number of recipients in control group and Figure 6 shows that of the treatment group.¹⁶ The black bars in each figure represent data on recipients before ARRA, and gray bars represent data in post-ARRA period. As discussed in the empirical model section, we obtain the number of recipients before ARRA by taking the average of those in years 2007-2009, and we calculate the number of recipients after ARRA effect by taking the average of those in years 2014-2016.

Number of Recipients in Control Group

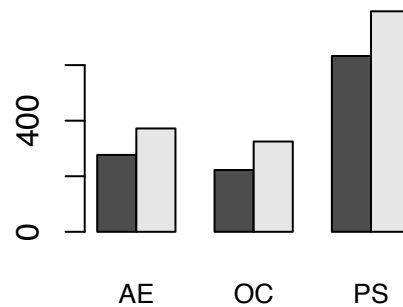


Figure 5: Number of Recipients of Fields in Control Group¹⁷

Number of Recipients in Treatment Group

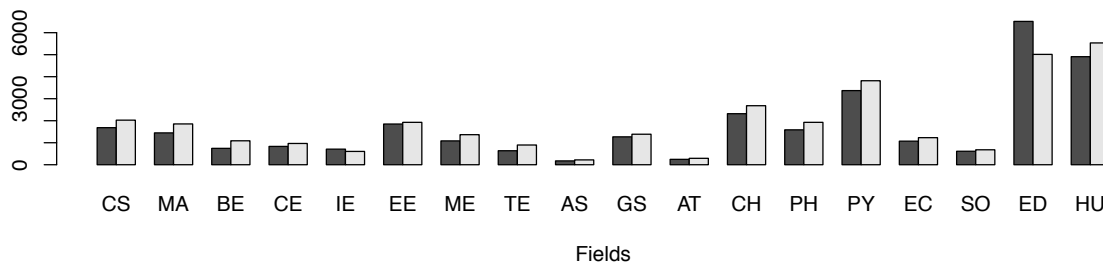


Figure 6: Number of Recipients of Fields in Treatment Group¹⁸

¹⁶ Note that the Y axes in Figure 5 and 6 have different ranges.

¹⁷ AE: Aeronautical Engineering, OC: Oceanography, PS: Political Science.

¹⁸ CS: Computer Sciences, MA: Mathematics, BE: Bioengineering and Biomedical Engineering, CE: Chemical Engineering, IE: Civil Engineering, EE: Electrical Engineering, ME: Mechanical Engineering, TE: Metallurgy and

In the second set of control and treatment groups, I match the *Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions* and the *Survey of Earned Doctorates (SED)*. This method provides me with 20 fields, in which 8 are in control group and 12 are in treatment groups. Figure 7 and 8 show the number of recipients in control and treatment groups under this setting. Similarly, black bars illustrate data for before ARRA and gray bars illustrate that of post-ARRA. The timing of ARRA affect is also the same as the previous setting.

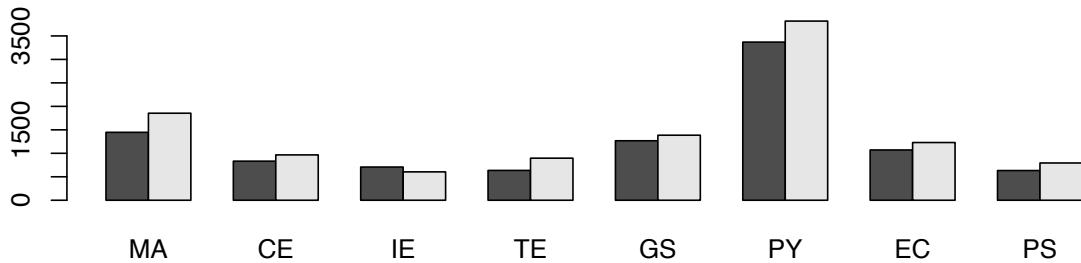


Figure 7: Number of Recipients of Fields in Control Group²⁹

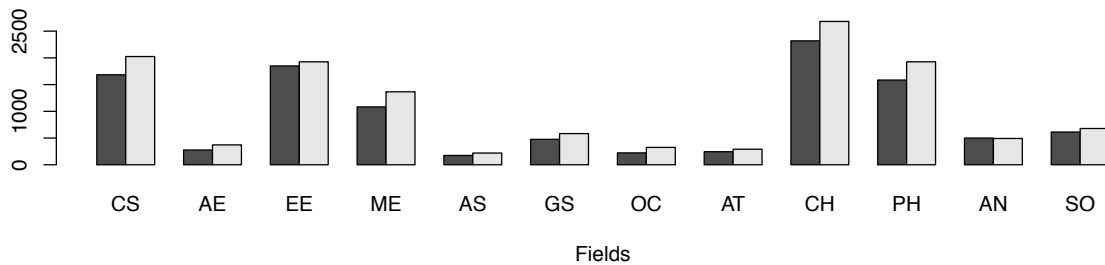


Figure 8: Number of Recipients of Fields in Treatment Group²

B. Independent Variables: Since this paper applies a difference-in-difference approach, the regression (same as the previous one)

$$Y = \beta_0 + \beta_1 *ARRA + \beta_2 *ARRA_affect + \beta_3 *ARRA *ARRA_affect + \varepsilon$$

independent variables include a dummy *ARRA* for before and after ARRA period, another dummy *ARRA_affect* for control and treatment groups, and an interaction term that gives us the difference-in-difference estimation. The selection of timing of ARRA effect, control and treatment groups were discussed in previous sections and shown in Figure 5,6,7 and 8. The level of research funding is not directed shown in the regression because rather we separate the control and treatment groups based on this data. However, the direct effect of level of funding will be tested with another regression and thus I include it in the summary statistics.

C. Summary: Table 1 presents summary statistics for the main variables used in the paper. In Setting 1, we observe a large difference in the number of recipients between the

Materials Engineering, AS: Atmospheric Sciences, GS: Agricultural Sciences, AT: Astronomy, CH: Chemistry, PH: Physics, PY: Psychology, EC: Economics, SO: Sociology, ED: Education, HU: Humanities.

¹⁹ An: Anthropology. All other abbreviations in Figure 7 and 8 are identical to those in Figure 5 and 6.

control and treatment groups. The average amount of science and engineering doctorate recipients in those fields in the treatment group, climbing from 1724 to 1861 during the ARRA period, is roughly 4 times that of in the control group. The level of funding also shows an increase during the ARRA period, from 534 million to 701 million dollars. Setting 2 yields similar patterns, in terms of the level of funding, before and after ARRA. As for the dependent variable, unlike setting 1, setting 2 produces similar number of recipients in the control and treatment groups; nevertheless, it also demonstrates an increase of this number during ARRA in both

Table 1: Summary Statistics

	Mean	SD	Min.	Median	Max.
<i>Dependent Variables: Number of Recipients</i>					
<i>Setting 1:</i>					
Control group1: before ARRA	377.2	222.88	222.3	276.7	632.7
Control group1: after ARRA	496.9	258.09	325	372	793.7
Treatment group1: before ARRA	1724.7	1664.19	174	1175	6512.3
Treatment group1: after ARRA	1861.2	1520.4	219.3	1375.5	5534
<i>Setting 2:</i>					
Control group2: before ARRA	1245.2	909.11	632.7	951.7	3368.3
Control group2: after ARRA	1443.2	1035.17	604.7	1097.8	3816.3
Treatment group2: before ARRA	918.8	753.95	174	556.2	2318.3
Treatment group2: after ARRA	1073.9	861.35	219.3	631	2681.7
<i>Independent Variables:</i>					
<i>Level of Funding (of all fields) in Setting 1:</i>					
Before ARRA (in 2008)	534144	340466.5	63348	386973	1210739
After ARRA (in 2012)	700975	453784.8	68428	551343	1603171
<i>Level of Funding (of all fields) in Setting 2:</i>					
Before ARRA (in 2008)	278460	297220	11040	158798	1089703
After ARRA (in 2012)	332438	340729	10093	217212	1258315

Notes: Dollars in thousands for Independent Variables. Refer "Setting 1" and "Setting 2" to the Empirical Model section.

groups. One noticeable set of numbers is the standard deviation (SD). This column shows that the variation of data is large, which is due to the fact that the number of observations in both settings are small- setting 1 incorporates 21 fields and setting 2 incorporates 20 fields. This is the main limitation of our datasets since the NSF did not record data on funding for smaller and more detailed subfields.

5. Results

(1) Main Regression Results

A. *Setting 1*: Table 2 presents the results from two OLS regressions under setting 1. The dependent variable Y is the number of doctorate recipients. The first regression tests the effect of ARRA funding on the number of recipients; the coefficient of the interaction term, which is the difference-in-difference estimator, is 16.78, with a standard deviation equal to 1331.45. This implies that fields that are affected by ARRA money on average has 17 more doctorate recipients after ARRA was implemented. However, the variance is large and the confidence interval at 95 percent confidence level is (-2678.6036, 2712.159). The 16.78 difference in the outcome variable is not significant at any level tested in the regression. The second regression model test the same effect on percentage level. The coefficient for $ARRA*ARRA_affect$ shows that being in the treatment group actually decreases the number of doctorate recipients of this academic field by 16 percent. Nevertheless, this is also an insignificant result. Therefore, from the first setting we conclude that putting money into an academic field does not have a significant effect on the corresponding number of doctorate recipients in that field.

Table 2: *Setting 1 Regression Results*

Independent Variables	(1)	(2)
	Y	Log(Y)
Intercept	377.22 (871.64)	5.83*** (0.5)
ARRA (Timing of ARRA)	19.67 (1232.68)	0.3 (0.71)
ARRA_affect (Control & Treatment)	1347.52 (941.48)	1.24* (0.54)
ARRA*ARRA_affect (D-in-D estimator)	16.78 (1331.45)	-0.16 (0.76)
Observations	42	42
Mean of dependent variable	1599.3	6.974

Notes: Standard errors in parentheses.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

B. *Setting 2*: In table 3, using observations from the second setting of control and treatment groups, we present similar regressions to those in the last section. The first model shows that having a larger increase in research funding during ARRA decreases the number of doctorate recipients by 43 units, and the second model demonstrates that being in the treatment group increases the quantity of educational outcome by 4 percent. But, similar to the results obtained in setting 1, neither of these coefficients is statistically significant. Therefore, this setting produces the same answer to our research question as in setting 1: ARRA money

does not affect the number of science and engineering doctorate recipients in post-ARRA period.

Table 3: Setting 2 Regression Results

Independent Variables	Dependent Variable	
	(1) Y	(2) Log(Y)
Intercept	1245.25*** (310.13)	6.96*** (0.28)
ARRA (Timing of ARRA)	198 (438.59)	0.14 (0.4)
ARRA_affect (Control & Treatment)	-326.47 (400.38)	-0.5 (0.36)
ARRA*ARRA_affect (D-in-D estimator)	-42.89 (566.22)	0.04 (0.51)
Observations	40	40
Mean of dependent variable	1135.5	6.746

Notes: Standard errors in parentheses.

*** Significant at the 1 percent level.

(2) Other Baseline Results

To check for a direct effect of the level of funding on the number of recipients. I run two simple regressions where I regress the number of recipients on the log of research funding for all fields, combining control and treatment groups in both settings. As seen in table 4, the regression coefficient of the first setting shows that a one percent increase in research funding decreases the quantity of recipients by 159 units, which is again an insignificant result. However, the result from setting 2 gives us a significant coefficient of log of research funding, showing that a one percent increase in ARRA money in this case increase the quantity of recipients by 374 units, which is significant at the 1 percent level. This is the only occasion when the data shows a significant and positive correlation between funding and quantity of graduates. However, this correlation is not necessarily casual. One possible explanation for this is that the R^2 for this regression is only 0.4, which means that the regression only explains 40 percent of the variation in the dependent variable. There are other factors, not included in this model, that could potentially affect the number of recipients.

Table 4: Baseline Simple Regression Result

Dependent Variable: Number of recipients	(1) Setting 1	(2) Setting 2
Independent Variables		
Intercept	3625.5 (3899)	-3362.2*** (891.6)
Log of Research Funding	-159.2 (297.6)	374.4*** (74.5)
Observations	42	40
R ²	0.07	0.4
Mean of dependent variable	1543.5	1087.7

Notes: Standard errors in parentheses.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

(3) Robustness Check

One major assumption of the main difference-in-difference model was regarding how I determine the value of *ARRA* variable. Recalling from the previous discussion, in this main regression model:

$$Y = \beta_0 + \beta_1 * ARRA + \beta_2 * ARRA_affect + \beta_3 * ARRA * ARRA_affect + \varepsilon,$$

for *ARRA* equal to 1 (post-*ARRA*), I claimed the value of the dependent variable, number of recipients, to be the average of the quantity in 2014, 2015 and 2016; for *ARRA* equal to 0 (pre-*ARRA*), I set the value of the dependent variable to be the average of the numbers in 2007, 2008 and 2009. Since there are no available documented evidence stating the average number of years it took for students to graduate after receiving *ARRA* money, the assumption that I made allowed for possible distortion of the actual effect of research funding. Therefore, I performed a few robustness checks with difference combinations of time frames to increase the confidence of my result.

Table 5 presents the results of these robustness check. The first and fourth columns show the regression coefficients from the original regression set-up that I defined under setting 1 and 2. For Robust1 and Robust3, I set the value of dependent variable to the average quantity of science and engineering doctorate recipients in 2015 and 2016 when *ARRA* is equal to 1 and set that value to the average number of recipients in 2008 and 2009 when *ARRA* is equal to 0. For Robust2 and Robust4, I claim the value of dependent variable to the number of doctorate recipients in 2016 when *ARRA* is equal to 1 and claim that value to the number of recipients in 2009 when *ARRA* is equal to 0. Therefore, I am able to test the effect of *ARRA* funding through different criteria in both settings.

Table 5: Robustness Check Results

Dependent Variable: Number of Doctorate Recipients	Setting1			Setting2		
	(1)	(2)	(3)	(4)	(5)	(6)
	Original	Robust1	Robust2	Original	Robust3	Robust4
Intercept	377.22 (871.64)	385 (873.9)	399 -878.5	1245.25*** (310.13)	1258.94*** (312.22)	1280.88*** (316.10)
ARRA (Timing of ARRA)	19.67 (1232.6)	113.17 (1235)	94 (1242.3)	198 (438.59)	197.69 (441.54)	176.50 (447.03)
ARRA_affect (Control & Treatment)	1347.52* (941.48)	1345.4 (943.9)	1342.2 (948.85)	-326.47 (400.38)	-339.23 (403.07)	-362.54 (408.08)
ARRA*ARRA_affect (D-in-D estimator)	16.78 (1331.4)	34.92 (1335)	39.06 (1341.9)	-42.89 (566.22)	-38.77 (570.03)	-19.25 (577.12)
Observations	42	42	42	40	40	40
Mean of dep. variable	1599.3	1609.8	1613.2	1135.5	1142.6	1145.8

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
Notes: Standard errors in parentheses.

The robustness checks show similar pattern to the original results. The coefficients for the difference-in-difference estimator in the *Robust* regressions have almost the same values as the original regressions. They all lead to the same conclusion: the effect of ARRA research funding on the number of science and engineering doctorate recipients is not significantly significant. Therefore, evidence from robustness checks for both settings increase our confidence that the estimates reported above are capturing the actual effect of ARRA on educational outcomes in terms of the quantity of doctorate recipients.

6. Conclusion

This paper estimates the effect of federal research funding provided during ARRA on the number of corresponding science and engineering doctorate recipients. Using a difference-in-difference approach, I compare academic fields before and after ARRA under two different settings, each resulting from a distinct combination of datasets. The regression estimation shows that fields that were heavily affected by ARRA money did not experience an increase in educational outcomes in terms of the number of doctorate recipients received. This conclusion holds true in both settings of control and treatment groups that I applied and is confirmed with the robustness checks.

Under setting 1, the main regression results show that fields that are targeted by ARRA had a roughly 17 units increase in the number of recipients; however, the standard error is relatively big, at 1331.45. Setting 2 provides us with similar results, except that it shows a small decrease, by 43, in the quantity of recipients for fields targeted by ARRA; nevertheless, the standard error is 566.22, also making this a statistically insignificant result. The reason for this

large standard error, which is also the main limitation of this result, is that the observations provided by the data is quite small. The NSF survey on the number of recipients has data on more than 100 academic fields; however, the surveys concerning federal research funding only contains data for around 30 fields. Therefore, under both settings, I am unable to get more than 25 fields that have data on both sides of the regression equation. If a bigger and detailed dataset is later collected for future studies, we can possibly reduce this variation in the regression results.

That being said, while past literatures mostly focused on the changing patterns of higher education R&D funding and employment outcomes of PhD students who received federal funding, this paper is the first to document the magnitude of ARRA effect by looking at the quantity of doctorate recipients. It fills into the gap within previous literatures and adds to a more comprehensive understanding of the impact brought by putting federal money into university research. The evidence presented in this paper, though it concerns funding from the federal government, also informs the impact of additional state resources on state-level educational outcomes. Future research, through more detailed and well-collected data, could obtain broader results in studying the effect of government money.