The Dynamics of Academic Science Research Production

Y. J. Jeff Qiu*

January 14 2018

Abstract

The present paper investigates the impact of funding on academic scientific research output by incorporating experience dynamics into the production function model. The model is estimated using panel data constructed from NIH cancer research grant archives. The results show that temporary reductions in funding can generate long-lasting decreases in research output, and are likely to hit less experienced labs more. Furthermore, the paper provides evidence in support of the prevailing view in the scientific community that the current funding allocations may be suboptimal, while the new funding regime initiated by the NIH is likely to further the progress of science.

*Department of Economics, Yale University. email: yinjia.qiu@yale.edu
I am indebted to my advisors Steve Berry, Phil Haile, Michi Igami, and Olav Sorenson for their guidance and encouragement. I also thank Dan Ackerberg, Don Andrews, Pierre Azoulay, Giovanni Compiani, Donghyuk Kim, Ying Fan, Wayne Gao, Soheil Ghili, Masayuki Sawada, Fabian Schrey, Jaya Wen, and participants at the Yale IO prospectus workshop for their valuable comments and suggestions. All potential errors are mine.
1 Introduction

Academic scientific research is at the core of innovation. Literature on the economics of science has documented the significant contribution of science to economic growth.\(^1\) Due to the nature of scientific knowledge as a public good, the federal government plays a significant role in supporting academic science research through various funding agencies, such as the National Institute of Health (NIH). In 2016, the NIH budget was estimated to be $35 billion, and most of the funding went to academics conducting basic science research. Public policies that determine the size of research funding have vital consequences for academic scientific research, and often provoke debates on how funding should be allocated to maximize social welfare. In order to address policy questions such as the static and dynamic effects of funding reduction and funding redistribution on aggregate scientific research output, a model is necessary.

The present paper investigates the impact of funding on academic scientific research output, and provides parameters necessary for policy recommendations. I propose a model of research production that takes funding and experience as inputs.\(^2\) Recognizing that research experience is an essential dynamic input into scientific research and that scientists can only accumulate experience when they have been provided with sufficient funding, the present paper models experience as a discounted accumulation of funding. As a result, funding not only enters research production directly as an input but also indirectly via the experience channel. My approach is in contrast to the previous literature (e.g. Jacob and Lefgren (2011)), which has estimated the direct effect of funding but neglected its dynamic impact.

Parameters in the production function, such as the elasticity of funding, are difficult to estimate due to endogeneity — variable inputs are correlated with productivity unobserved by the econometrician. The present paper employs techniques from the production function literature (e.g. Olley and Pakes (1996), Blundel and Bond (1998, 2000), Ackerberg, Caves, and Fraser (2015)) by exploiting the dynamic panel structure to tackle endogeneity issues that arise in production function estimation.

The present paper focuses on academic cancer research labs. Using lab-year level panel data constructed from the NIH grant archive, the estimated model in this study demonstrates the importance of the dynamic impact of funding through the

---

\(^1\)See Stephan (1996 and 2012) for a complete survey of the literature.

\(^2\)Research output is measured as a sum of number of quality-weighted publications.
experience channel, and provides parameters that are necessary for offering policy recommendations on designing rules for funding allocation.

First, I employ the estimated model to study the impact of temporary funding reduction — a policy similar to the one proposed by the Trump administration\(^3\) — on research output over time. The results suggest that such a policy may cause a long-lasting reduction in output through the experience channel, and are likely to harm younger labs more significantly.

Second, I utilize the model estimates to investigate whether the current funding allocation rule is dynamically inefficient by not sufficiently investing in younger, less established labs. It is well documented in the literature that the age distribution of academic scientists is shifting to the right, and that most of the research funding is allocated to prominent scientists while less established scientists are underfunded (Blau and Weinberg (2017)). Scientists and policy makers have proposed to increase support for younger researchers, arguing that the current funding environment is hurting the future of science as young talents are unable to secure necessary funding to produce research output.\(^4\)

Despite these overwhelming arguments against the current dissimilar funding allocation between established scientists and young scientists, the inequality we observe in the data may not be surprising — established labs are more experienced, and thus the marginal product of funding is higher for these labs. This creates a feedback loop. Merton (1968) coined the term “the Matthew Effect” to describe this phenomenon in the sociology literature, where the more resources scientists already have, the more they obtain, and vice versa. However, due to diminishing marginal returns of research experience and its dynamic consequences, the current funding allocation may generate inefficiency. Using the estimated model, I compare the aggregate output generated from an allocation that imposes a funding cap with one generated from the current funding allocation, and show that the current funding allocation may be suboptimal.

The paper is organized as follows. Section 2 reviews related literature. Section 3 discusses the institution setup and data construction. Section 4 presents the model and estimation strategies. Section 5 shows the results. Section 6 illustrates the implications of the model through two counterfactual exercises. Section 7 concludes.

---

\(^3\)https://www.nytimes.com/2017/04/03/us/politics/trump-medical-research-funding-nih.html?_r=0

\(^4\)https://www.nytimes.com/2014/10/03/opinion/young-brilliant-and-underfunded.html
2 The Economics of Science

2.1 The Impact of Funding

The significant effect of scientific research on economic growth can be classified into two categories: the indirect benefit of research funding to the local economy (e.g., Weinberg et al (2014)), and technological progression due to advancements in basic science from academic research. Most works in the economics of science literature have focused on the latter channel and documented important impacts of scientific discoveries on innovation, and hence on long-term economic growth.

For example, using state-technology level panel data, Jaffe (1989) found evidence for a strong spillover effect of academic research to industry research output as measured by patent count, and showed that this effect was especially prominent in the areas of chemical and biomedical research. Adams (1990) examined the impact of scientific knowledge — measured by the stock of discounted publications from nine fields (e.g., biology and medicine) — on the growth in firms from 18 industries, and found significant contributions of knowledge to each firm’s total factor of productivity. However, the impact of scientific knowledge on the productivity of firms are not immediate. Mansfield (1991) surveyed 76 major American firms in seven manufacturing industries, and found that 11 percent of their products could not have been developed without findings in recent academic research.

From the above, it is apparent that a strong relationship exists between academic research and innovation, and since almost all academic scientific research is funded by the federal government in the US, it is important to estimate the effect of funding on research output. Previous works have attempted to estimate the elasticity of funding to research, but found mixed results. For example, Adams and Griliches (1998) found a statistically insignificant elasticity of funding at the university-field level analysis, and attributed the findings to measurement error in variable inputs. Jacob and Lefgren (2011) used administrative data from the NIH, and also found statistically insignificant effects of receiving an NIH grant on research output. Their approach relied on a regression discontinuity design in the cutoff of the priority score, an application score assigned by panels of referees to determine funding allocation. Hence, their finding only applies to those at the score cutoff point, and cannot be representative for all research production. Recently, using exogenous variations for NCAA tournaments, Tabakovic and Wollmann (2016) focused on a selected sample
of universities and reported a higher estimate of elasticity of funding. Azoulay, Graff-Zivin, Li, and Sampat (2017) linked patents from pharmaceutical and biomedical industries to citations of academic research output, and found that a 10-million-dollar increase in NIH funding is associated with an increase of 2.3 patents.

The present paper contributes to the literature by providing a model and estimation strategies in the production function literature, and I find that funding has an important impact on research output both statically and dynamically.

2.2 Learning and Forgetting

Scientific research production processes are intrinsically dynamic because research output is a function of experience, and experience can only be accumulated by doing research. The key observation is that a research lab’s experience develops like human capital. It accumulates through various investments, and depreciates due to worker turnover and/or depreciated relevance of past knowledge to current research topics. Theoretical models in the literature usually include the experience dynamic channel, but empirical works generally neglect it.

In the theoretical model of Levin and Stephan (1991), a model of human capital is used that includes both knowledge accumulation and depreciation. In their theoretical model, they considered a Cobb-Douglas research production function that takes knowledge — measured by a fraction of the researcher’s stock of publications — as a variable input for producing research output. It is clear that the human capital accumulation and depreciation rates have important consequences on research activity in their model, but their empirical analysis disregards this dynamic.

The human capital model studied in the economics of science literature is analogous to the research development models developed by Griliches (1979). For example, Adams and Griliches (1998) modeled discounted accumulation of research capital as a variable input in production. However, their estimation procedure assumed that the depreciation rate is known.

Furthermore, similar models can be found in the learning-by-doing and organizational forgetting literature. In particular, recognizing the importance of production experience as a dynamic input for aircraft production, Benkard (2000) developed production models, and estimated both the learning rate (the impact of experience (or human capital) on production) and depreciation rate of experience.
The present paper models research experience as a dynamic input for the production function similar to Griliches (1979), Adams and Griliches (1998), and Benkard (2000), and develops a procedure to estimate both the impact of research experience on production and the depreciation rate of experience.

2.3 Life Cycle Models

The present paper is also related to one branch of the literature that addresses research productivity of scientists over a life cycle. Works in this literature have utilized a human capital investment framework to model the research production of academic scientists over their life cycles. For example, Levin and Stephan (1991) modeled that researchers enjoy utility from consumption and research output, and that their objective is to allocate time to research in order to maximize utility, which consists of income and research output. They found that research activity declined as researchers aged, both theoretically and empirically. Jones (2010) proposed a theory where the accumulation of knowledge over a generation leads to a higher requirement for updated knowledge over time.

These models suggest a rise in productivity in the early careers of researchers and a decline as researchers age, and the patterns are investigated in empirical studies (e.g., Diamond (1986), Stephan and Levin (1993), Jones and Weinberg (2011)). A general pattern that the productivity of researchers declines over a life cycle is documented, but depends on the field of studies.\footnote{For example, Stephan and Levin (1993) found no evidence that particle physicists exhibit the productivity pattern predicted in the human capital model.}

2.4 Funding Allocation

How to best allocate research funding is one of the most important questions in the literature and the public policy arena. Current debates focus on how funding should be allocated to benefit scientific progress in terms of identifying promising research projects and helping younger scientists. The current NIH funding allocation mechanism is criticized for its inability to prioritize good research proposals, its lack of statistical rigor, as well as its bias in gender, race, and how established a lab is (e.g., Li and Agha (2015), Li (2017)).
Currently, the average age of first-time recipients of the R01 grant, the most important grant in all life sciences, is 42 (the average was 32 in 1981), and younger scientists are underfunded (Harris (2014)). It has been documented in several studies that the age distribution of scientists is shifting upward (Blau and Weinberg (2017)), where the life cycle model suggests research output to decline. Furthermore, funding inequality is severe, with 10% of the grant recipients receiving 40% of the research funding. In addition, studies have shown diminishing returns of funding to research output and argued that the current funding allocation may be inefficient (e.g., Fortin and Currie (2013) and Cook et al (2015)).

Alternative allocation rules that have been proposed share a common theme — to promote young researchers (e.g. Fang and Casdella (2016), and Fang, Bowen, Casadella (2016)). The present paper contributes supporting evidence for the recent debate that research funding has a stronger impact on younger research labs because of underlying experience dynamics, and that the current funding allocation may be inefficient.

3 Data

The present paper makes use of detailed grant level data from the NIH. This dataset includes information of grants from 1985 to the present, and contains funding amount, Principal Investigator (PI) information, the review section to which the grant was sent, budget period, abstract, etc. Only successful applicants are observed in this dataset. Grants are divided into different activities. For example, the most common grant is R01, which is awarded to PIs for conducting research projects in an area that represents their specific interest. It is well known that most life science labs depend on R01 grants to stay active in academic research (Stephan 2012). There are grants that are provided on the institutional level to establish research centers, such as C06 and P30 grants. Furthermore, grants are separated into different types such as new grants, competing renewal, and noncompeting renewal. “New grant” and “competing renewal” applications must be examined in peer review sections to determine the winners.\(^7\)

\(^6\)https://www.nih.gov/about-nih/who-we-are/nih-director/statements/new-nih-approach-grant-funding-aimed-optimizing-stewardship-taxpayer-dollars

\(^7\)See Stephan (2012) and Azoulay, Zivin, and Manso (2012) for in-depth surveys of the funding agency.
In life sciences, an academic research lab is managed by a PI, typically a professor at a university, who receives research funding from governmental agencies such as the NIH for operation. The funding received by PIs is spent on acquiring input resources such as capital and labor (postdocs, graduate students, and other scientists), which the labs employ to produce research output (publications, patents, clinical studies, etc).\footnote{In effect, I use the terms PI and lab interchangeably throughout the present paper.} Scientific research, especially in the biological sciences, can only be conducted with sufficient funding.

Although we can in principle study all disciplines in the life sciences, the current paper focuses on one particular biological science: academic cancer research. I follow approximately 4000 cancer research labs identified by recipients of R01 grants from the National Cancer Institute (NCI),\footnote{The budget for NCI was $5 billion in 2016.} the largest institute center at the NIH, from 2000 to 2014.\footnote{I collect all grants that these cancer research labs received from the NIH in a given year and aggregate the funding amount to the lab-year level. These include grants received from different institutional centers.} I further restrict my sample to labs that received their first R01 grant after 2000. Thus, in effect, only the most relevant cancer research labs are considered in my sample.

There are several reasons why I have chosen to base my analysis on cancer research. First, by focusing on one particular science, the analysis is more tractable, and comparisons between different labs are more valid, as different types of life sciences may have different research production processes. My choice thereby mitigates potential confoundedness between two different production functions.\footnote{This, of course, assumes that cancer research labs with different disease specializations have the same production function. We could relax this assumption by disaggregating the data to the lab-disease-year level, and estimate the production function for different diseases separately. Appendix C shows an example on how to define specialization of labs. However, this disaggregation opens up an avenue of problems. First, it is difficult to associate labs with certain disease specialities. Second, even if we are able to identify labs with certain disease specializations, since most labs have multiple and different specializations over time, we need further assumptions to associate these labs with the proper disease speciality. Third, there are many diseases within the cancer category, so we face the additional problem of the choice of granularity of disease aggregation. Finally, the identification strategy utilized in this paper crucially relies on the panel structure, which is data demanding. Therefore, a finer level of aggregation implies that less data is available.} Second, NIH funding accounts for the major share of funding to cancer research labs. Because of the prominence of funding from the NIH, it is approximately valid for me to assume that my sample of labs is representative of all labs in this sector, thus avoiding any selection bias. By extension, I assume that the NIH is the sole source of funding, and
that labs are unable to operate without its funding.\textsuperscript{12} Finally, cancer is one of the deadliest diseases that threaten public health. Understanding the academic cancer research production function can provide policy guidance that are most effective in helping to maximize cancer research output.

A typical grant lasts for four to five years and allocates funds to labs in 12-month budget periods. Hence, it is reasonable to define the unit of observation at the lab-year level. I construct the funding amount $F_{it}$ of lab $i$ at time $t$ by aggregating the budget allocated to lab $i$ in year $t$.\textsuperscript{13} The NIH also contains data on all publications that cited the grant number, and I measure the research output of a lab at time $t$ by a measure that represents the quality and contribution of publications written by the lab in year $t + 1$ that cite the grant, which is calculated as a weighted sum.\textsuperscript{14} Formally, I define the output of $i$ at time $t$ as

$$Y_{it} = \sum_{p \in P_{i,t+1}} \frac{\varphi_p}{\#(p)},$$

where $P_{i,t+1}$ denotes the set of papers published by lab $i$ in year $t + 1$, $\varphi_p$ is paper $p$’s research output as measured by its quality, and $\#(p)$ denotes the number of labs involved in paper $p$. The division of $\varphi_p$ by $\#(p)$ captures the research output proration, which is often used to address publications with multiple coauthors, and is common in the economics of science literature. In contrast to the measure used in the literature, a slight nuance in (1) is that the paper output is prorated by the number of labs rather than the number of coauthors, because our analysis is at the lab level.\textsuperscript{15}

If $\varphi_p = 1$ for all $p$, then $Y_{it}$ is the unweighted measure of research output, which simply aggregates the number of papers the lab published. This unweighted measure only captures the quantity of publication and ignores the quality of the papers.

\textsuperscript{12}This neglects funding from the private sector and grants from the Howard Hughes Medical Institute (HHMI). See Azoulay, Graff Zivin, and Manso (2011) for the difference between NIH and HHMI grants.

\textsuperscript{13}In general, grant level data display the total cost of each grant, which contains direct and indirect costs. Indirect costs refer to funding amounts used to pay the grant recipient, i.e., overhead of the home institution of the PI. Direct costs refer to the actual amount of funding received by the PI. The funding variable $F_{it}$ defined here represents the total cost. This implicitly assumes a constant proportion of overhead paid by the PIs to different institutions.

\textsuperscript{14}Life science publications are usually much shorter than social science ones; also, the review time of publications is generally much shorter for life sciences than for social sciences.

\textsuperscript{15}This type of measurement is commonly used in the economics of science literature and the patent and innovation literature.
alternative definition of output would be to measure $\phi_p$ by the number of citations paper $p$ has. However, citations of publications take time, which may contaminate the more recent data, as we only have data from 2000 to 2014. In this paper, I employ the impact factor of the journal in which paper $p$ is published as a measure of $\phi_p$, such that the resulting weighted output measure would not suffer from inconsistencies of quality measures over time. This measure, although unable to distinguish between the quality of two papers published in the same journal, is adequate in capturing quality differentiations across different journals.\textsuperscript{16}

4 A Model of Academic Science Production

4.1 Production Function

This section presents a model of research production and lays out the assumptions and their dynamic implications. This paper focuses only on the modeling and estimation of production. For an example of a dynamic model studying the optimal dynamic funding allocation problem using results implied by the model presented here, see Qiu (2017).

I focus only on established academic science labs, and it is assumed that the labs must be recipients of NCI R01 grants prior to production and that the labs maintain their R01 grant recipient status during the period of production. I do not consider the pre-R01 grant production of PIs for several reasons. First, academic scientists typically spend years of postdoctoral training after their doctoral studies to work under another PI before their first R01 grant, and while they work in a lab, they work under guidance of the PI, the owner of the lab. Second, academic life scientists in the US will need to win an R01 grant in order to carry out their own research agenda. That is, upon receiving an R01 grant, PIs act similarly to a CEO of a firm, managing resources (funding) and giving directions to lab members to produce

\textsuperscript{16}Alternatively, we can measure output such as the number of patents that have cited the papers associated with a particular NIH grant. This measure is proposed and carefully constructed in Azoulay et al (2017). However, it takes a longer time to include all patents that cite publications associated with a grant. For instance, about 40% of NIH grants awarded in 1991–1995 are associated with at least one patent 10 years after their approval (Azoulay et al (2017)). Using this patent output measure would potentially yield many zero outputs in the lab-year panel structure adopted in the present paper; therefore, this measure is unsuitable for the Cobb-Douglas production function model considered in the present paper.
research output. By focusing on PIs, we effectively treat each lab as an independent firm that takes in variable inputs to produce research output. For an analysis that studies the transition from postdocs to PIs, see Kim, Qiu, and Sawada (2017) and Lerchenmueller and Sorenson (2017).

A research lab’s production function is assumed to follow the Cobb-Douglas form, and takes funding and experience as variable inputs to produce research output:

\[
\ln Y_{it} = \alpha_0 + \alpha_1 \ln F_{it} + \alpha_2 \ln E_{it} + \alpha_3' X_{it} + \omega_{it} + \varepsilon_{it},
\]

where the subscript \( it \) indicates lab \( i \) at time \( t \), output \( Y_{it} \) is defined by (1), \( F_{it} \) and \( E_{it} \) are funding and experience variable inputs for the lab respectively, \( X_{it} \) is a vector of the lab’s observed characteristics such as time trend, school fixed effects and age, \( \omega_{it} \) is the productivity of the lab observed by the lab and the NIH but not observed by the econometrician, and \( \varepsilon_{it} \) is the measurement error of output. The key parameters of interest are \( \alpha_1 \), the elasticity of funding, and \( \alpha_2 \), the elasticity of experience, a parameter that describes learning. Learning rate typically refers to the percentage reduction in input when experience doubles. Here, I differ from the use of terminology, and learning rate refers to the percentage increase in research output when experience doubles, that is, \((2^{\alpha_2} - 1) \times 100\%\). The model (2) constrains the parameters of the production function to be fixed across labs and over time.\(^{17}\) The model studied in this paper extends naturally to more flexible functional forms, such as the trans-log production function.\(^{18}\)

There are a few differences between (2) and the standard setup of production function in the production function literature. First, I do not include labor input as funding input already encapsulates it. Possible measures of labor input may include the number of postdocs, Ph.D. students, undergraduate students, and technicians in the research lab. Generally, we do not have this information in the NIH dataset, but

\(^{17}\)Labs that do not produce any papers over the entire sample period are dropped. The output measure, \( Y_{it} \), is transformed by incrementing it by one in order to allow for labs that produce zero output for certain years. This transformation, used by many researchers (e.g., Tabakovic and Wollmann (2016)), is a normalization allowing for minimum output for labs to remain in the sample.

\(^{18}\)The output definition (1) coupled with the model (2) implicitly assumes that publications at time \( t + 1 \) are associated with funding and experience at time \( t \). As alluded to earlier, publications in life sciences are typically short and requires a shorter review process than social sciences. Nonetheless, we may try different lag definitions. For instance, we can associate publications at time \( t + 2 \) with funding and experience at time \( t \). However, this assumes zero contribution of funding at \( t + 1 \) to papers published at \( t + 2 \).
we could approximate it using the number of coauthors in the publications of a given PI. In academic science in the United States, PIs need to provide financial support for postdocs, Ph.D. students, and other labor input using their grants. This practice may be very different for academic science in Asia, for example, where postdocs salaries and Ph.D. student stipends are paid by the PIs’ home institutions in some cases. However, given that we only focus on NIH grant recipients, this assumption is innocuous. Second, we have experience, \( E_{it} \), as a variable input. It can be viewed as a state variable that determines research output. This variable is the key dynamic aspect in research production. It is included to capture the important feature of research, especially in life science, that research output is a function of extensive trial and error. This input is similar to the R&D capital input in the knowledge production function introduced by Griliches (1979).

### 4.2 Experience

I model the experience component of (2) similar to the knowledge models in Griliches (1979) and Levin and Stephen (1991), as well as the learning models in the learning-by-doing and organizational forgetting literature, e.g., Benkard (2000).\(^{19}\)

The key feature of experience is that it captures the accumulation of relevant research capital over time. In traditional learning models, experience is defined as

\[
E_{it} = E_{i,t-1} + R_{i,t-1}, \text{ with } E_{i\tau_i} = \theta_i, \tag{3}
\]

where \( R_{i,t-1} \) is a measure of research capital of lab \( i \) at time \( t \), \( \tau_i \) indicates the entry year of lab \( i \), and \( \theta_i \) is the initial condition. In this straightforward model, \( E_{it} \) simply measures the accumulation of research capital since entry.\(^{20}\) This simple specification assumes that the experience of a given lab gains \( R_{i,t-1} \) in each period, conditional on whether they are in the market, and that experience does not depreciate.

The deterministic model described above assumes no depreciation of experience. It is documented that organization forgetting is an important feature of dynamic and capital accumulation dynamics, (e.g, Pakes and Schankerman (1979), Levin and Stephan (1991), and Benkard (2000)). Therefore, I model the experience dynamic as

---

\(^{19}\)Knowledge plays the same role as experience in the production process, and I use the experience model terminology throughout this paper. I use the terms experience and knowledge interchangeably.

\(^{20}\)A lab’s entry year is measured by the year in which it receives its first R01 grant.
follows:
\[ E_{it} = \delta E_{i,t-1} + R_{i,t-1}, \text{ with } E_{ir_i} = \theta_i, \]  
\hspace{1cm} (4)

where \( \delta \) is the rate of experience depreciation. In this model, a lab can improve its experience only by acquiring research capital, and experience depreciates at a constant rate \( \delta \). The constant rate of depreciation is specified in an attempt to capture worker turnover and depreciation of relevant knowledge, etc. The model (4) refers to the experience process in a general learning model. The difference between traditional learning models and general learning models is that the former assume \( \delta = 1 \) while \( \delta \) in the latter is unknown and will be estimated.

While the experience model (4) is similar to the knowledge process in human capital models in the economics of science literature (e.g. Levin and Stephan (1991)), they are conceptually different. In the literature, knowledge process refers to scientific knowledge accumulation of a researcher. In the present paper, a lab is modeled as a firm, and the experience process refers to research capital accumulation of a lab (e.g. Griliches (1979)). The experience process (4) encapsulates both scientific knowledge of the PI and research experiences of lab members. As a result, depreciation rate \( \delta \) in the literature refers to the depreciation of past scientific knowledge of a researcher, but depreciation rate \( \delta \) in (4) refers to depreciation of relevant knowledge of the PI as well as employment effects on a micro-level such as lab member turnover and training of new lab members. Since scientific research is labor intensive and turnover rate is high, we can expect a higher depreciation rate for model (4) than for knowledge processes in the literature.

One candidate for \( R_{i,t-1} \) is \( F_{i,t-1} \), the funding amount \( i \) received in period \( t - 1 \). In this specification, experience improves by how much funding the lab received in the previous period. When \( \theta_i = 0 \), \( E_{it} = \sum_{j=1}^{t-1} \delta^{j-1} F_{i,t-j} \), which is a measure of \( i \)'s discounted accumulation of funding. In general, \( R_{i,t-1} \) could be measured as the number of grants won, number of papers published, number of top publications, etc.

Although similar, the research production model studied in the present paper, described by (2) and (4), is conceptually different from the learning model in the learning-by-doing literature. First, in the learning models studied in the literature (e.g. Benkard (2000)), experience refers to discounted cumulative past output. Therefore, in these models, experience only accumulates if the firm produces output. Scientific research production, however, is a function of trial and error. Scientists typically only publish successful research, and thus failed experiments are often left unac-
knowledged, even though failed experiments are a crucial component of and pathway to experience accumulation.

For instance, suppose we take some measure of research output, say the number of publications, as a measure of research capital in (4), then the contribution of failed experiments to the experience dynamic is immediately omitted, despite its fundamental importance to scientific research. Therefore, research capital measured by output would substantially understate the actual experience. In order to model this special feature of the scientific experience dynamic, discounted accumulation of funding is a reasonable measure of experience input.\textsuperscript{21}

Second, learning rate in the learning-by-doing literature refers to the percentage reduction in labor input when experience doubles. In the learning model presented here, learning rate refers to the percentage increase in research output when experience doubles.

Even though there are several candidates for $R_{it}$, they serve the same fundamental idea, that experience accumulates and depreciates. This paper focuses on discounted accumulation of funding as a measure of experience as the benchmark model for two reasons. First, as stated earlier, research capital measured by funding can better capture the trial and error aspect of science experience dynamic. Second, one drawback of the output definition (1) is that it may be sensitive to different time lag definitions. For example, (1) with production function (2) assumes away the impact of funding at $t-2$ on research output at $t$. Experience dynamic measured as discounted cumulative funding, however, mitigates this problem because past funding is also an input to current research output. Hence, past funding has an indirect effect on current research output. Nevertheless, experience as measured by the cumulative number of publications is considered for comparison purposes.

Before I discuss the estimation procedures, it is necessary to address a few points. First, it is reasonable to consider a production function with more than one experience input. For example, we may have one that measures discounted accumulation of funding and one that measures discounted accumulation of number of publications. In this setup, we will need to estimate two discount factors and two learning coefficients. This estimation will suffer from multicollinearity, because funding and the number of

\textsuperscript{21}This measure omits the contribution of failed grant applications to experience. However, we are interested in the effect of experience on research output and not the effect of experience on writing grant proposal. Therefore, the contribution of failed grant applications may not understate the actual experience for research production.
papers are highly correlated, and there is no good instrument that shifts the variation of one experience but not the others (Griliches (1979)). Second, I acknowledge that there are multiple types of experience inputs that are essential determinants to research output. The model described by (2) and (4) is a simplification of the complex production function.

4.3 Productivity Structure

Direct OLS estimation of (2) is problematic if the variable inputs can be determined with knowledge of $\omega_{it}$. As a result, $F_{it}$ and $E_{it}$ are correlated with $\omega_{it}$, which is the source of endogeneity. Identification and estimation of the production function parameters depend crucially on the structure of the productivity $\omega_{it}$.

One simple solution to the endogeneity problem is to assume $\omega_{it} = \omega_i$, that is, a lab specific fixed effect. This allows us to estimate the parameters through standard individual fixed effect estimators, such as first differencing. However, the assumption that the unobserved productivity of the labs stays constant over time is strong, especially with a long panel like the data used in the present paper. Furthermore, estimating the production function under fixed effect estimators is shown to face potential problems of severe biases (e.g., Griliches and Hausman (1986)) and simply do not work well in practice (e.g., Ackerberg, Benkard, Berry, and Pakes (2007)).

One alternative solution is to use instrumental variables that are correlated with the variable inputs but do not determine output and are not correlated with $\omega_{it}$ and $\varepsilon_{it}$. Recent works estimating production function with instrumental variables include Benkard (2000), Bloom, Schankerman, and Van Reenen (2013), Tabakovic and Wollmann (2016), and Azoulay, Graff-Zivin, Li, and Sampat (2017). In general, instruments are difficult to obtain. However, due to the long panel structure, variation of NIH budget over time could be used as instruments for the inputs in (2).

Another solution is to impose a Markov structure along with timing assumptions on the unobserved productivity $\omega_{it}$ (e.g., Olley and Pakes (1996), Levinsohn and Petrin (2000), and Ackerberg, Caves, and Fraser (2015)). That is, the productivity process follows

$$\omega_{it} = \Psi(\omega_{i,t-1}) + \xi_{it},$$

where $\Psi(\cdot)$ is a nonparametric function, and $\xi_{it}$ is an innovation shock centered at
the origin with the orthogonality condition

$$\xi_{it} \perp \mathcal{I}_{i,t-1}$$

(6)

where $\mathcal{I}_{i,t-1}$ is the information set of lab $i$ up to period $t-1$, a set of characteristics observable to the NIH and the lab. Condition (6) is a common timing assumption used in the literature and yields the key identification condition. To be exact, the timing assumptions on the production process implied by (5) and (6) are the following: labs observe $\mathcal{I}_{i,t-1}$ and apply for research funding with proposed amount $F_{it}$ at the end of period $t-1$, then the NIH observes $\mathcal{I}_{i,t-1}$ for all $i \in \mathcal{N}_t$ where $\mathcal{N}_t$ is a set of all labs applying for the grant, and makes decisions about whether to fund the labs. Labs that are funded move on to period $t$, then the innovation shock $\xi_{it}$ is realized, and the labs at $t$ produce research output $Y_{it}$ according to technology (2).

The timing assumptions and structures (5) and (6) serve as an important identification condition in the literature.\textsuperscript{22} This identification condition relies on the knowledge of innovation shock $\xi_{it}$ at the time decisions are made. It is common that uncertainties enter research production. For example, the success of novel research topics can go in either direction (Stephan (1996)), implying that the decision makers do not have perfect information of $\xi_{it}$ at $t-1$. And the above assumption implies that the best predictor for $\omega_{it}$ is $\Psi(\omega_{i,t-1})$.

### 4.4 Identification and Estimation

Estimation of the production function (2) suffers from classic endogeneity issues because the variable inputs are chosen based on $\omega_{it}$, which is unobserved.\textsuperscript{23} The present paper employs a version of the productivity structure (5) and the timing assumption (6), now standard in the practice of the production function estimation literature, to tackle the endogeneity problem. In general, there are two approaches in the production function literature: the proxy variable approach, and the dynamic panel approach.

In the proxy variable approach in the production function estimation literature (e.g., Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves,\textsuperscript{22}This type of timing assumptions have also been used in the demand estimation literature (e.g. Lee (2013)). See Ackerberg and Hahn (2015) for recent results on nonparametric identification of production function using these timing assumptions.

\textsuperscript{23}Note that $\varepsilon_{it}$ is exogenous.
and Fraser (2015)), researchers often employ a control function for $\omega_{it}$, typically using intermediate input or investment, together with a Markov productivity structure, to identify and estimate the primitives of the production function. Specifically, this approach assumes the existence of a function, say

$$M_{it} = \Phi_t(\omega_{it}, F_{it}, E_{it})$$

(7)

where $M_{it}$ is an observed variable excluded from the production function (2), and $\Phi_t(\cdot)$ is an unknown function of productivity, funding, and experience. Furthermore, $\Phi_t(\cdot)$ is assumed to be strictly increasing in $\omega_{it}$, and cannot have additional unobserved heterogeneity. Due to the monotonicity assumption of $\Phi_t(\cdot)$ in $\omega_{it}$, the unobserved productivity can be written as $\omega_{it} = \Phi_t^{-1}(M_{it}, F_{it}, E_{it})$; together with the structures (5) and (6), we can identify and estimate the parameters of the production function.

Candidates for $M_{it}$ could be, for example, intermediate input and grant application score. One caveat is that these variables are rarely available to the public. Nevertheless, some researchers have access to the NIH’s administrative data (e.g. Azoulay, Graff-Zivin, Li, and Sampat (2017), and Li (2017)); thus, it is possible to measure $M_{it}$ as the application score. However, even with assistance from application score data, additional assumptions are needed to formulate an analysis suitable for the proxy variable approach. For instance, PIs may have multiple applications in the same year, and hence multiple application scores are available for that year. Coupled with the fact that the scores are assigned by different study sections, with possibility different tastes for grant proposals, the aggregation of application scores to the lab-year level is nontrivial.

In the dynamic panel approach, (e.g. Blundel and Bond (1998, 2000)), an additional restriction is imposed on (5). Specifically, (5) is strengthened to follow a first order autoregressive (AR1) structure

$$\omega_{it} = \rho \omega_{i,t-1} + \xi_{it},$$

(8)

where $\rho$ captures the persistency of the productivity process.\textsuperscript{24} This method is well-studied in the econometrics of dynamic panel literature (e.g., Anderson and Hsiao (1982), and Arrellano and Bond (1991)).

\textsuperscript{24}In general, a linear productivity structure is needed. For instance, we can have AR2 productivity structure, i.e. $\omega_{it} = \rho_1 \omega_{i,t-1} + \rho_2 \omega_{i,t-2} + \xi_{it}$.\n
By $\rho$-differencing of (2), conditions (8) and (6) yield the following moment conditions

$$
\mathbb{E} [\xi_{it} + \varepsilon_{it} - \rho \varepsilon_{i,t-1} | I_{i,t-1}] = 0 \\
\mathbb{E} [(\ln Y_{it} - \alpha_0 - \alpha_1 \ln F_{it} - \alpha_2 \ln E_{it} - \alpha_3' X_{it}) - \\
\rho (\ln Y_{i,t-1} - \alpha_0 - \alpha_1 \ln F_{i,t-1} - \alpha_2 \ln E_{i,t-1} - \alpha_3' X_{i,t-1}) | I_{i,t-1}] = 0.
$$

(9)

To identify the parameters of interest, we need funding variables from at least three different periods. The underlying identification condition requires that the past funding a lab received (conditional on observables) is not a perfect predictor for future funding of the lab. The sources of variation come from different types of smaller grants a lab can receive from the NIH, application luck (due to the taste of the study section), and variations in the NIH budget.

Both approaches have their own strengths and weaknesses. The proxy variable approach requires more assumptions on the existence of a control function for the unobserved productivity, and restriction to unobserved heterogeneity, but it allows the unobserved Markov productivity process to be nonparametric. The dynamic panel approach does not require the existence of a control function, but imposes a stronger assumption on the productivity process, which has to be linear. Nevertheless, both approaches rely crucially on the same timing assumption (6). See Ackerberg (2016) for a detailed discussion of both approaches. Due to data limitations, I employ the dynamic panel approach to identify and estimate the parameters in (2) by exploiting the panel structure.

Parameters in (2) can be estimated with (9) using Generalized Method of Moments (GMM). The set of instruments for the moment condition (9) is all the information of lab $i$ up to and including period $t - 1$. In general, we can use all observables in the information set of lab $i$ at $t - 1$, $I_{i,t-1}$, including any lags in funding and output prior to $t - 1$. We may also include $F_{it}$ in $I_{i,t-1}$. This assumes that $F_{it}$ is not correlated with the innovation shock, which is a stronger timing assumption. Generally, the inclusion of $F_{it}$ into $I_{i,t-1}$ would yield more precise estimates, but this assumption would be violated if lab $i$ or the NIH is aware of extra information about $\xi_{it}$ before proposing for funding $F_{it}$. Other instrumental variables include supply and demand shifters such as budget for the NIH, budget of different type of cancers, death and
incidence rates of different cancers, and any lags in these variables.

One nice feature of the dynamic panel model is that the moment conditions (9) yield over-identification restrictions, which allows us to test the validity of different instruments via Sargan-Hansen’s J-test. This, in turn, tests the validity of different timing assumptions.

Furthermore, the framework developed in the present paper extends naturally to different definitions of time aggregation. For instance, we can define time bi-annually instead of annually as in our baseline model. The panel and productivity structure would remain valid. See Appendix B for a robustness analysis using this level of aggregation.

Note that it is assumed that the NIH makes the funding decision for period \( t \) after observing \( I_{i,t-1} \) but before the realization of the innovation shock \( \xi_{it} \). This timing assumption guarantees that

\[
E[\xi_{it}|I_{i,t-1}, \text{exit}_{it} = 0] = E[\xi_{it}|I_{i,t-1}] = 0,
\]

where \( \text{exit}_{it} \) indicates that lab \( i \) exits at period \( t \). Without this assumption, an exit model or a funding allocation model would be needed to correct for sample selection.\(^{25}\)

5 Results

This section presents results of the production estimates under three different productivity structures: exogeneity, fixed effect, and AR1. Results of the production estimates under exogeneity and fixed effect productivity structures are shown in Section 5.1. They are presented mainly for comparison purposes. Results under AR1 productivity structure are shown in Section 5.2. I estimate the production function with two output measures: sum of number of quality-weighted publications (weighted output) where quality is measured by journal impact score, and sum of number of publications (unweighted output), to assess the differences of the impact in funding on the quantity and quality of output.

\(^{25}\)This argument is used in the literature (e.g., Levinsohn and Petrin (2003), Doraszelski and Jaumandreu (2013), and Ackerberg, Caves, and Fraser (2015)).
5.1 Descriptive Statistics

Table 1 shows estimates for the production function that uses weighted output. Two deterministic experience measures are used: cumulative number of publications and cumulative amount of funding.\textsuperscript{26} Columns (i) and (ii) of table 1 present OLS estimates while columns (iii) and (iv) present fixed effect estimates.\textsuperscript{27} In all cases, the age of the labs, defined by number of years since their first R01, and school ranking by the total funding received from the NCI, are added as controls.\textsuperscript{28} Time fixed effects are included for the OLS regressions. Table 2 shows the estimates of the same regression but uses unweighted output.

Results from columns (i) and (ii) in Tables 1 and 2 are biased because variable inputs are endogenously determined. The correlation between log weighted output and log funding ranges from 0.315 to 0.381, and the implied learning rate is about 54\% under the cumulative number of publications experience measure and 10.3\% under the cumulative amount of funding experience measure. On the other hand, the correlation between log unweighted output and log funding ranges from about 0.181 to 0.226, which is significantly lower than the estimates using weighted measure of output. The implied learning rate is also higher, which is 29\% under the cumulative number of publications experience measure and 7\% under the cumulative amount of funding experience measure. In addition, the negative coefficient of school ranking found indicates that higher tier schools is positively correlated with output, which is expected. Furthermore, a negative coefficient of lab age is found.

Columns (iii) and (iv) in Tables 1 and 2 present the fixed effect version of research production function under weighted and unweighted output. If one is willing to assume that productivity is constant over time, the results would yield consistent and unbiased estimates of the parameters for the production function. Under the fixed effect model, the elasticity of funding is about 0.187 and 0.117 under weighted and unweighted measure of (1) respectively. These estimates are significantly lower than the estimates from OLS (columns (i) and (ii) Tables 1 and 2), suggesting the crucial endogeneity problem with variable inputs: higher unobserved productivity is associated with higher variable inputs; therefore, OLS estimates of the elasticity of funding is upward biased.

\textsuperscript{26}Note that (3) with \( \delta = 1 \) refers to traditional learning.
\textsuperscript{27}A first difference estimator is employed.
\textsuperscript{28}The lower the number, the higher the rank.
One problem presented in column (iii) in Tables 1 and 2 is that the coefficient of experience when measured in cumulative number of publications makes little sense. This suggests that defining experience as an accumulated number of papers is problematic. This is not surprising, however. First, the number of papers a lab produces is quite stable, and fixed effect estimators require differencing of the variable inputs. Therefore, the differencing results in insufficient variation of the inputs. Second, using number of papers as research capital is not a good measure because this measure significantly underrepresents the contribution of failed experiments to research experience. Third, without controlling for cumulative amount of funding, the output measure defined in (1) is sensitive to the timing of lags. Finally, if experience input is subject to measurement error, the fixed effect estimator would be downward biased (Griliches and Hausman 1986). A fixed effect model with experience specified as the cumulative amount of funding, however, yields robust and reasonable results.

Overall, the OLS and FE estimates illustrate that the correlation between output and funding is higher when the output is weighted.

5.2 Production Function Estimates

Table 3 shows the dynamic panel estimates, where experience is modeled as discounted cumulative funding. The results presented here are estimated using two-step GMM with moment conditions (9). Instruments include \( X_{it}, F_{i,t-1}, F_{i,t-2}, F_{i,t-3}, Y_{i,t-2}, NIH \) budget, and death and incidence rates of cancer. In addition, columns (ii) and (iv) include \( F_{it} \) as one of the instruments while columns (i) and (iii) exclude it. For the NIH budget as well as death and incidence rates of cancer, time lags are also used. Columns (i) and (ii) present results under the traditional learning model (\( \delta = 1 \)), while columns (iii) and (iv) present results under the general learning model. Only

---

29For example, consider the following model \( y_{it} = x_{it} \beta + \omega_i + \varepsilon_{it} \), where \( \omega_i \) is the fixed effect and \( \varepsilon_{it} \) is an exogenous shock. Suppose we observe \( z_{it} = x_{it} + \eta_{it} \) instead of \( x_{it} \), where \( \eta_{it} \) is an iid measurement error. It can be shown that fixed effect estimators are biased. For example, a first difference fixed effect estimator can be written as \( \hat{\beta} = \beta \left( 1 - 2\sigma^2_{\eta} / \text{var} (\Delta z_{it}) \right) \) where \( \Delta z_{it} = z_{it} - z_{i,t-1} \) and \( \sigma^2_{\eta} \) is the variance of \( \eta_{it} \). If \( \Delta z_{it} \) has little variation, then the estimated coefficient could be negative. In the production model, measuring experience as the cumulative number of publications is a poor estimate of experience input because it only accounts for the contribution of success, therefore it is subject to severe measurement error problem. Furthermore, since the number of papers a given lab publishes does not vary much over time, there is little variation in the change of experience with this measure, yielding the unreasonable results in the fixed effect model shown in column (v) in Tables 1 and 2.

30Appendix A illustrates a construction of budget instrumental variables.
weighted output are considered in this section. Heterostedasticity Autocorrelation Consistent (HAC) standard errors are reported. Furthermore, I use Sargan-Hansen’s J-test to test the validity of the moments (9) implied by the dynamic panel model. The J-test cannot reject the null hypothesis that the moments in (9) are valid for all models estimated in Table 3 at the 10% level.

The following estimates were obtained in models that include $F_{it}$ as an instrument. Under the traditional learning model (column (ii)), the estimated elasticity of funding is 0.204, and the implied learning rate is 20%. Under the general learning model (column (iv)), the estimated elasticity of funding is 0.22, the implied learning rate is 17%, and the estimated depreciation rate is 0.561. Here, all the key production function parameter estimates are statistically significant. In all the models presented in Table 3, the autocorrelation coefficient $\rho$ is about 0.9, showing persistency of the productivity process. The coefficients of age and school ranking are similar to those presented in Tables 1 and 2, but some are not statistically significant. In all of the regressions in Tables 1 - 3, research output declines as the lab age increases, supporting the life-cycle hypothesis. Note, however, that the age defined in the present paper is the number of years since the first R01 grant. Therefore, it does not capture the real effects of the PI’s age. Results that exclude $F_{it}$ as an instrument yield similar results, but with bigger standard errors.

The estimated elasticities of funding are higher than those according to the fixed effect models and lower than those using OLS. These estimates are comparable to those found in Tabakovic and Wollmann (2016) and higher than those found in the previous literature (e.g. Adams and Griliches (1998) and Jacob and Lefgren (2011)).

Focusing on column (iv), the estimated depreciation rate is 0.561 with a standard error of 0.183. This implies that 56.1% of stock experience existing from the beginning of the year remains relevant at the end of the year. The estimate may seem high. However, as pointed out earlier, the experience process contains the scientific knowledge of the PI, research experience of the lab members, and other forms of research capital. Thus, the depreciation rate does not merely measure the depreciation of scientific knowledge of the PI.

The high depreciation rate can be justified by the organizational structure of academic research labs and the nature of scientific research. First, lab members are typically Ph.D. students and postdocs. They are not permanent lab members, and may move between labs through different stages of their career or quit because they
find non-academic career options to be more attractive.\textsuperscript{31} In addition, new lab members are not perfect substitutes for those who left since training takes time. Hence, the high depreciation rate can be explained by the high turnover rate of lab members. Second, scientific research production requires numerous scientific experiments. Some of the experimental designs are unique and may not be easily transferable. Lastly, as pointed out by Stephan (2012), certain fields progress quickly, so that scientists (both the PI and lab members) may not be able to keep up with the pace at which the discipline is changing over time.\textsuperscript{32}

6 Counterfactual Experiments

My production model implies that research funding is crucial to production both statically and dynamically. Statically, funding is a variable input to production. Dynamically, funding enters the production through the experience channel; the dynamic impact of funding depends on the learning rate, depreciation rate, and level of experience. This section uses the parameter estimates from column (iv) in Table 3 to evaluate the implications of the funding dynamic through counterfactual funding allocation scenarios.

6.1 Funding Reduction and Expansion

In a world in fear of the decline of research funding — for instance, the Trump administration’s proposal for a 20\% funding cut — it is important to evaluate the impact of different funding policies to research output. To illustrate this type of scenario, I hypothetically reduce funding to all existing labs at a uniform rate over two years. Recall that our research production model is

\begin{align*}
Y_{it} &= A_{it} (F_{it})^{\alpha_1} (E_{it})^{\alpha_2} \\
E_{it} &= \delta E_{i,t-1} + F_{i,t-1}.
\end{align*}

\textsuperscript{31}See Preston (2004) for a review of occupational exit from scientific careers.

\textsuperscript{32}As a comparison, the annual experience depreciation rate is estimated to be 0.61 in the aircraft manufacture industry (e.g. Benkard (2000)). This estimate is quite high considering that the aircraft production process is relatively stable.
In the first year (period $t$) the funding cut does not enter experience, so this uniform cut translates to a $100(1 - \tau)^{\alpha_1 - 1}\%$ reduction of output, where $\tau = 0.2$. A uniform 20% cut would imply a 4.79% reduction of research output with a standard error of 0.41% in the initial year.\(^{33}\)

In the second year (period $t + 1$), however, the funding cut from the first year affects experience, and the implied experience is $E_{i,t+1}^\tau = (1 - \tau)F_{it} + \delta E_{it}$. The percentage reduction in research output in period $t + 1$ can thus calculated to be

$$\frac{Y_{i,t+1}^\tau}{Y_{i,t+1}} - 1 = (1 - \tau)^{\alpha_1} \left( \frac{E_{i,t+1}^\tau}{E_{i,t+1}} \right)^{\alpha_2} - 1$$

$$= (1 - \tau)^{\alpha_1} \left( \frac{(1 - \tau)F_{it} + \delta E_{it}}{F_{it} + \delta E_{it}} \right)^{\alpha_2} - 1$$

$$= (1 - \tau)^{\alpha_1} \left( \frac{(1 - \tau)R_{it} + \delta}{R_{it} + \delta} \right)^{\alpha_2} - 1 =: \Delta_{t+1}(\tau, R_{it})$$

where $R_{it} = F_{it}/E_{it}$, which is the funding-to-experience ratio. The percentage change of research output in period $t + 1$ depends on the elasticity of funding, the learning rate, the depreciation rate, and the funding-to-experience ratio. One can verify that $\Delta_{t+1}(\tau, R_{it})$ decreases with an increasing funding-to-experience ratio. Since younger labs have little experience at the beginning of the research production process, the ratio $R_{it}$ is higher; hence, the percentage reduction in research output is higher. To validate this observation, I construct the experience measure using estimates from column (iv) in Table 3, and a simple regression of funding-to-experience ratio on the age of the lab reveals a negative relationship between the funding-to-experience ratio and age. To illustrate the heterogeneous dynamic impact of funding, I compare the computed $\Delta_{t+1}(\tau, R_{it})$ with a high funding-to-experience ratio $R_{H} = 0.98$, which is the median ratio corresponding to two-year-old labs, and a low funding-to-experience ratio $R_{L} = 0.38$, which is the median ratio corresponding to two-year-old labs. A 20% cut of research funding would further reduce research output by 7.68% in the second period with a standard error of 0.88% for $R_{H}$ (younger labs). On the other hand, the same 20% cut of research funding would reduce research output by 6.59% in the second period with a standard error of .59% for $R_{L}$ (more established labs).\(^{34}\)

\[^{33}\]Standard errors of the counterfactuals are constructed using the Delta method.

\[^{34}\]Likewise, I repeat the same counterfactual exercise but hypothetically raise research funding by 20% uniformly to all labs. I find that a 20% expansion would yield a 4.09% increase in research output with a standard error of 0.37% in the first period. In the second period, since the funding
To further illustrate the dynamic impact of funding, we will compare two different funding scenarios. First, consider a scenario in which labs receive the same amount of funding — the amount received in 2013 — over 10 years. In this case, the labs accumulate experience according to (4), their productivities are held fixed, and they produce output via technology (2). Now consider a policy in which the funding of all labs is cut by 20% from the level of 2013 for the first four years, and the labs receive the original amount of funding for the remaining six years. Furthermore, I assume no entry and exit in this counterfactual exercise. Figure 1 shows the percentage of output reduction compared to the control under this policy with the two different experience levels over time. Overall, the percentage reduction of output peaks at year four, the last year of the cut, with a 8.67% reduction in total output. This translates to about 2000 units of quality research output reduction, equivalent to about 70 top ranked journal articles.

Figure 1 presents evidence that the policy has heterogeneous impacts as less experienced labs suffer more due to diminishing return of funding and experience. Furthermore, Figure 1 also demonstrates that temporary funding reductions have long lasting effects on output. Even when funding returns to the original level at year five, research output is still reduced by 4.5%. This counterfactual exercise demonstrates the importance of the funding dynamic to output, because funding reduction in the initial period slows down experience accumulation, which in turn affects output in the subsequent periods, even when the regime of funding reduction ends. The output reduction is mitigated as time progresses in this counterfactual design, however, due to the depreciation of experience and diminishing marginal returns of experience.

The results from this counterfactual design show that funding reduction would bump has an impact on experience. Research output increases by 6.9% with a standard error of 0.92% for $R_H$ (younger labs), and 5.94% with a standard error of 0.6% for $R_L$ (more established labs). The labs’ initial states (funding, experience, and productivity) are set as their observed states in 2013. The same peer-review system would likely remain active, and the allocation of funding to each lab may be different due to a budget cut. Some existing labs may not survive under a tighter budget. Therefore, the counterfactual analysis of 20% budget cut on the existing model with only the production function may be unable to account for such a situation. The same is true with a funding expansion, because such policy is likely to induce more entry. A more comprehensive model is needed to account for general effects of funding cut and expansion.

This plots $100\left(\frac{\sum_i Y_{it}^{\text{cut}}}{\sum_i Y_{it}^{\text{original}}} - 1\right)$ % over time. $Y_{it}^{\text{cut}}$ indicates the output lab $i$ would have produced under the funding cut regime in period $t$, and $Y_{it}^{\text{original}}$ indicates the output lab $i$ would have produced under the 2013 funding allocation.
have a greater impact on younger labs. The reverse is also true.

6.2 Funding Redistribution

“Because scientific discovery is inherently unpredictable, there are reasons to believe that supporting more researchers working on a diversity of biomedical problems, rather than concentrating resources in a smaller number of labs, might maximize the number of important discoveries that can emerge from the science we support” — NIH director Francis Collins

The current NIH funding allocation shows severe inequality — 10% of grant recipients receive more than 40% of the total research grants. Evidence from Fortin and Currie (2013), Cook et al (2015), and the present paper have shown diminishing marginal returns of research funding to output, which may imply inefficiency in the current funding allocation. Recently, NIH director Francis Collins proposed a new NIH funding allocation rule that caps the amount of funding a lab can receive. The goal of this reform is to redistribute the funds from established scientists to early- and mid-career scientists as a way to increase research output.\(^{38}\)

In the second counterfactual exercise, I hypothetically impose a $1 million funding cap to all labs and redistribute the excess funding to other labs. I assume that labs receive the same amount of funding they received in 2013 over 6 years; labs accumulate experience according to (4), their productivities are held fixed, and they produce according to technology (2). In my sample, there are 2821 labs receiving $1.3 billion in 2013 and 230 labs receiving $1 million. Together, they received $325 million. A one million dollar cap would yield an excess of about $123 million, and the excess is distributed to the remaining labs. I compare the aggregate outputs over time under the observed funding allocation and three alternative allocations. The first allocation distributes the excess amount evenly to the rest of the labs. The second allocation distributes the excess amount evenly to less experienced labs. The third allocation distributes the excess amount evenly to high productivity labs.\(^{39}\)

---

\(^{38}\)Blogpost: https://www.nih.gov/about-nih/who-we-are/nih-director/statements/new-nih-approach-grant-funding-aimed-optimizing-stewardship-taxpayer-dollars


\(^{39}\)Productivities are taken to be the residuals. This implicitly assumes no measurement error \(\varepsilon_{it}\).
Figure 2 illustrates the percentage increase in aggregate research outputs under the observed allocation and three alternative allocations. All three alternatives outperform the observed allocation. Allocations one and two give similar results. Although allocation two underperforms compared to allocation one in the first year because less experienced labs have lower marginal products of funding, it outperforms allocation one as time progresses, because less experienced labs benefit more from funding expansion dynamically. Allocation three dominates the first two allocations because the marginal product is higher for high productivity labs. Note, however, that funding support for the established labs may provide exceptional training for numerous postdocs and Ph.D. students. The effect of funding on the development of future scientists is difficult to quantify and neglected in the counterfactual analysis. This, in turn, underestimates the impact of funding reduction on established labs.

These simple redistribution rules are undoubtedly suboptimal. First, they assume that the NIH cares only about aggregate output; in reality, the NIH may prioritize certain fields. Second, in order to design an optimal funding allocation to maximize objective, say aggregate output, the NIH needs to consider dynamic trade-offs between experience and productivity, because labs with less experience have higher marginal returns to funding dynamically. Despite these concerns, the counterfactual exercises demonstrate that the current allocation rule may be suboptimal. This is a consequence from the diminishing marginal returns of funding and the experience dynamics.

7 Conclusion

This paper studies the impact of funding on academic research output by incorporating experience dynamics into the production function. The estimation follows methodologies developed in the production function literature by exploiting dynamic panel structure and timing assumptions. Using the estimated model, two counterfactual exercises provide important implications on how funding affects research production both statically and dynamically. Importantly, these experiments show that temporary funding reductions may have long-lasting effects on output through the

\[ \text{This plots } 100 \left( \frac{\sum_{i} Y_{it}^{\text{redistribution } j}}{\sum_{i} Y_{it}^{\text{original}}} - 1 \right) \% \text{ for allocation } j \in \{1, 2, 3\} \text{ over time. } Y_{it}^{\text{redistribution } j} \text{ indicates the output lab } i \text{ would have produced under the redistributed allocation } j \text{ at period } t, \text{ and } Y_{it}^{\text{original}} \text{ indicates the output lab } i \text{ would have produced under the 2013 funding allocation.} \]
experience channel, and is likely to hit less experienced labs more significantly. Furthermore, these exercises provide evidence in support of the view in the scientific community that the current funding allocation may be suboptimal, while the new funding regime initiated by the NIH director, Francis Collin, is likely to accomplish advances in science.

References


8 Appendix A: Tables

Table 1: Regression Baseline Weighted

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Fixed Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
</tr>
<tr>
<td><strong>Dependent variable:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>log weighted output</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td>0.315***</td>
<td>0.381***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Experience: # publications</td>
<td>0.541***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>Experience: funding</td>
<td>0.141***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>School Ranking</td>
<td>-0.030***</td>
<td>-0.040***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.113***</td>
<td>-0.025***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Observations</td>
<td>25,157</td>
<td>25,157</td>
</tr>
<tr>
<td>R²</td>
<td>0.248</td>
<td>0.121</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.248</td>
<td>0.121</td>
</tr>
</tbody>
</table>

*Note:* *p<0.1; **p<0.05; ***p<0.01
Table 2: Regression Baseline Unweighted

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Fixed Effect</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
<td>(iii)</td>
</tr>
<tr>
<td>Funding</td>
<td>0.181***</td>
<td>0.226***</td>
<td>0.117***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Experience: # publications</td>
<td>0.364***</td>
<td>−0.470***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Experience: funding</td>
<td></td>
<td>0.094***</td>
<td>0.081***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>School Ranking</td>
<td>−0.006**</td>
<td>−0.013***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>−0.080***</td>
<td>−0.021***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>25,157</td>
<td>25,157</td>
<td>21,069</td>
</tr>
<tr>
<td>R²</td>
<td>0.296</td>
<td>0.134</td>
<td>0.117</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.296</td>
<td>0.133</td>
<td>0.117</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>log weighted output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
</tr>
<tr>
<td>Funding</td>
<td>0.224***</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.261**</td>
</tr>
<tr>
<td></td>
<td>(0.115)</td>
</tr>
<tr>
<td>School Ranking</td>
<td>−0.085**</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
</tr>
<tr>
<td>Age</td>
<td>−0.019</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.901***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
</tr>
</tbody>
</table>

| Observations       | 17,208   | 17,208   | 13,649    | 13,649    |
| GMM J-test p-value | 0.236    | 0.234    | 0.274     | 0.167     |
| Is $F_{it} \in I_{i,t-1}$? | no      | yes      | no        | yes       |

Note: $^*$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01
Table 4: Robustness Check: Two Years Aggregated Panel

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>FE</th>
<th>DP</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
<td>(iii)</td>
<td>(iv)</td>
</tr>
<tr>
<td>Funding</td>
<td>0.509***</td>
<td>0.301***</td>
<td>0.234**</td>
<td>0.287***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.019)</td>
<td>(0.106)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.060***</td>
<td>0.155***</td>
<td>0.197*</td>
<td>0.164**</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.118)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>School Ranking</td>
<td>−0.037***</td>
<td>−0.063</td>
<td>−0.133</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.074)</td>
<td>(0.081)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0005</td>
<td>0.037</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.047)</td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td>δ</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.276*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.148)</td>
</tr>
<tr>
<td>ρ</td>
<td></td>
<td></td>
<td></td>
<td>0.901***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.023)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.886***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.027)</td>
</tr>
<tr>
<td>Observations</td>
<td>10,794</td>
<td>8,756</td>
<td>5,374</td>
<td>3,281</td>
</tr>
<tr>
<td>GMM J-test p-value</td>
<td></td>
<td>0.111</td>
<td>0.536</td>
<td></td>
</tr>
<tr>
<td>Is F_{it} ∈ I_{i,t-1}?</td>
<td>NA</td>
<td>NA</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
Figure 1: Impact of the Trump Cut by Experience

![Graph showing the impact of the Trump Cut by Experience. The graph plots percentage of output change against year. There are three lines representing overall, high experience, and low experience categories. The graph shows a decrease in output change from the initial year, with the high experience category showing the least decrease, followed by the overall category, and then the low experience category.]
Figure 2: Funding Redistribution

- Allocation 1: evenly distributed
- Allocation 2: low experience
- Allocation 3: high productivity
9 Appendix B: Robustness

Production function estimation suffers from endogeneity issues, because labs that are more productive are likely to receive more funding and gain even more experience. To tackle the endogeneity issue, the present paper utilizes methodologies developed in the production function literature. The identification arguments rely crucially on the panel structure and timing assumptions.

The present paper focuses on the lab-year level panel data structure. I thereby assume that papers published in year $t+1$ is associated with input variables of the lab in year $t$. The decision for this level of aggregation was made based on the life science publication cycle and the 12-month budget period of the NIH.

The framework developed in the present paper extends naturally to different time aggregation. In this appendix, I aggregate the data to the lab-time level where the time definition is defined biennially (every two years), and perform robustness analyses. To illustrate this framework, let $\tau$ represent a particular year in time. We define

$$ F_{i,t+j} = F_{i,\tau+2j} + F_{i,\tau+2j+1} $$
$$ Y_{i,t+j} = \sum_{p \in P_{i,\tau+2j+1} \cup P_{i,\tau+2j+2}} \frac{\varphi_p}{\#(p)} $$

for $j = 0, 1, 2, ...$

Experience, $E_{it}$, is accumulated according to (4). The subscript $t$ is measured in a unit that increments by one every two years. I only consider observations every two years. For instance, for a lab that is in the sample for 7 years, I construct three data points from the first 6 years. The information from the seventh year is not included. One drawback with this aggregation is that we have a lot less observations.

The same regression analysis is performed using this level of aggregation, and the results are presented in Table 4. Only weighted output are considered. Columns (i) and (ii) of Table 4 present the OLS and Fixed Effect results respectively. Similar patterns are shown, that the correlation between log funding to log output with the OLS estimate is 0.509, and it is higher than the Fixed Effect estimate of 0.301, indicating the presence endogeneity in the production function. However, these estimates are higher than the lab-year level aggregation.

Columns (iii) and (iv) of Table 4 presents the dynamic panel results from tradi-
tional learning \((\delta = 1)\) and general learning \((\delta \text{ not restricted})\) respectively. These results are comparable to those presented in Table 3. The depreciation rate estimate is 0.276. This may seems much lower than the depreciation rate estimate from the lab-year level analysis, which is 0.561. But recall that the depreciation rate at this level of aggregation implies that 27.6\% of stock experience existing from the beginning of year \(t\) remains relevant at the end of year \(t + 2\). Using our estimate from the lab-year level analysis, the implied depreciation rate at the biennial level would be \(0.561^2 \approx 0.315\), which is close to the estimate using the biennial aggregation.

While this aggregation produce similar results, there are a few differences worth pointing out. First, the instrument used in the dynamic panel model excludes \(F_{it}\). This is due to the structure of the aggregation, because the innovation shock \(\xi_{it}\) includes shocks to output in year \(\tau\) and \(\tau + 1\), or papers published in year \(\tau + 1\) and \(\tau + 2\). However, since \(F_{it}\) includes funding at year \(\tau\) and \(\tau + 1\) by definition, it would not be orthogonal to the shock \(\xi_{it}\). Second, the number of observations using dynamic panel drops significantly, because the dynamic panel model, especially the dynamic panel model, requires at least three lags for identification. The aggregation used in this appendix reduces the number of lags available, which greatly reduces the number of observations for the analysis. This is a crucial drawback at this level of aggregation.
10 Appendix C: Data Construction

10.1 Lab Specialization

I identify the specialization of labs by the types of cancer studies they produce. Let \( \mathcal{C} \) be a set of cancer types considered in this paper. This set is constructed using the broad spending categories defined by the NIH.\(^{41}\) The set \( \mathcal{C} \) has 10 elements. The first 9 elements represent 9 major cancer types (brain cancer, breast cancer, cancer genomics, cervical cancer, colorectal cancer, liver cancer, lung cancer, pancreatic cancer, and prostate cancer) and the last element represents all other cancer types.

I utilize the MeSH code dataset to identify the specialization of labs. The MeSH code, or medical subject heading, is a comprehensive controlled vocabulary for the purpose of indexing journal articles in the life sciences. Each life science publication is attached with several MeSH codes, and the codes categorize the types of diseases studied. For example, C04.588.180 indicates breast neoplasms and C04.588.614.195 indicates brain neoplasms.\(^{42}\)

There are several ways to categorize the labs. One way is to create a vector of binary variables \( c_i = [c_{i,1}, \ldots, c_{i,10}]' \) where \( c_{i,k} = 1 \) if lab \( i \) has published at least one paper that studies cancer type \( k \) and 0 otherwise. Another way is to simply count the number of papers that study cancer type \( k \) lab \( i \) has published. In addition, we can put a subscript \( t \) in \( c_i \) to indicate time. In effect, we allow the specializations of the labs to change over time.

10.2 Budget Construction

This paper uses several types of NIH budgets as instruments for variable inputs in the production function. The overall NIH fiscal budget is publicly available. Similarly, the NCI fiscal budget is publicly available.\(^{43}\) Budgets on different cancer types are available from the NIH, but only for a limited period of time (2008 to present). Since the data used in this paper is based on year 2000 to 2014, the numbers published from the NIH cannot be used. Instead, I consider a proxy for cancer types budgets based on MeSH codes.

\(^{41}\)https://report.nih.gov/categorical_spending.aspx
\(^{42}\)See https://meshb.nlm.nih.gov/search for more information on MeSH code.
\(^{43}\)https://www.cancer.gov/about-nci/budget
In order to construct a proxy for budgets for different cancer types, I aggregate the amount of funding to the year-cancer type level. Specifically, I define budget for cancer type \( k \in C \) in period \( t \) as

\[
B_{kt} = \sum_{i \in N_t} F_{it} w_{ikt}
\]

where \( N_t \) is the set of cancer labs, \( F_{it} \) is the amount of funding \( i \) receives in period \( t \), and \( w_{ikt} \) is lab \( i \)'s weight assigned to cancer type \( c \) in year \( t \). Two measures of weights can be used. The first measure simply assigns 1 to the weight if the lab publishes at least one paper that studies cancer type \( k \) in period \( t \). So \( w_{ikt} = 1\{c_{ikt} > 0\} \). The second measure considers a weighted version. Specifically, I define \( w_{ikt} = c_{ikt} / \sum_{k \in C} c_{ikt} \) where \( c_{ikt} \) is the number of papers \( i \) published on cancer type \( k \) in period \( t \).