How DOEs Government Funding Fuel Scientists?

Eugenie Dugoua* Todd Gerarden† Kyle Myers‡ Jacquelyn Pless§

November 19, 2025

Proposal: NBER Energy Markets and the Macroeconomy Conference

Abstract

We study how direct government funding for research and development shapes scientific and economic activity in the energy sector. Our empirical strategy uses quasi-experimental variation in technology-specific funding "windfalls" that emerge when Congress appropriates more or less than what the Department of Energy (DOE) requests. We link this funding to data on energy-related dissertations and patents to quantify the effect of government funding on science and innovation outcomes. In ongoing work, we are extending our research design to analyze broader economic outcomes. We are in the process of linking DOE research funding to firm-level microdata from Compustat based on their technological similarity. We will then analyze the impact of DOE funding on firm-level outcomes including research and development spending, productivity, and output. Finally, we will develop a model to assess the aggregate economic effects of DOE funding for research and development.

*Department of Geography and Environment, London School of Economics (e.dugoua@lse.ac.uk)
†Cornell Dyson School of Applied Economics and Management (gerarden@cornell.edu)
‡Harvard Business School (kmyers@hbs.edu)
§MIT Sloan School of Management (jpless@mit.edu)

1 Introduction

Technological change has long been recognized as central to economic growth (Aghion and Howitt 1992; Romer 1990). Private markets tend to under-provide innovative activity because the full social value of discoveries and inventions is not typically captured by their creators. This phenomenon is most frequently appreciated in the context of firms' innovation decisions, as imperfect appropriability can dampen the incentives for firms to invest in research and development (R&D). But it applies to the entire pipeline from basic research to applied innovation—training of new scientists, production of new ideas, and the commercialization decisions of firms—all of which may be subject to distorted incentives.

Yet most policies focus on providing incentives to firms to invest in innovative activity rather than creating scientific human capital (Van Reenen 2021). Governments frequently offer direct grants and tax credits that reduce the cost of investing in R&D, and there is growing evidence that these incentives enhance innovative activity.[1] An implicit assumption behind these policies is that more innovation by firms may increase the demand for inventors and lead to more individuals engaging in innovative activity. However, if the supply of inventors is inelastic, these demand-side policies may just increase the wages of high-skilled workers (Goolsbee 1998). The supply of inventors could be inelastic in the short run due to frictions in higher education (Romer 2000). Human capital policies may therefore be the most direct mechanism for increasing the supply of inventors (Bloom, Van Reenen, and Williams 2019; Van Reenen 2021). After all, reaching the frontier of knowledge and producing new ideas typically relies upon significant education and training. But providing financial incentives for entry into the market for ideas may have limited benefits if the supply of researchers and potential inventors is also inelastic. Understanding how government funding shapes both the training of new scientists and the downstream production and commercialization of new technologies is therefore an important open question.

In this paper, we trace out how government funding for research affects scientific and economics outcomes in the energy sector. We do so by estimating the effect of funding from the U.S. Department of Energy (DOE) allocated to different energy technologies on related dissertations, patents, and firm-level measures of innovative activity and economic performance. Quantifying the

1. See Bloom, Van Reenen, and Williams (2019) for a review of this literature.

impacts of government funding on the supply of inventors is empirically challenging for several reasons. First, standard measures of the number of inventors—such as authors of patents—do not capture the supply of potential inventors, as individuals may engage in innovative activity without yet successfully innovating. Second, even when innovation occurs, it may manifest both through scientific output from academia and applied outcomes from firms, each of which may respond differently to funding. Identifying the causal effects of government funding is complicated by the fact that the government may allocate funding based on unobservable factors that independently influence the production of scientists and science as well as the innovation and performance of firms active in those technologies.

We overcome the first of these two challenges by building a novel data set of PhD dissertations produced by U.S. higher education institutions. We find that both the level and share of energy-related science, technology, engineering, and mathematics (STEM) dissertations increased significantly from the 1980s to today, with a lot of the growth starting in the late 2000s. Dissertations across all energy technologies increased, including "dirty" energy, but the number of "clean" energy-related dissertations experienced much faster growth. We complement these data with technology-specific patent records to capture innovative output, allowing us to examine how funding affects not only the entry of new researchers but also the ideas they generate.

We overcome the second challenge by using quasi-experimental variation in government funding. We frame our empirical model as a production function estimation problem in the spirit of Olley and Pakes (1996) to account for unobserved productivity at the technology-year level that is likely to bias estimates of the relationship between DOE funding and the production of scientists, scientific output, and innovation. To estimate the effect of DOE funding on dissertations and patents, we use variation in technology-specific funding "windfalls" from when Congress appropriates more or less than the amounts requested by the DOE. These windfalls arise because Congressional appropriations committee members have personal or local preferences or are influenced by lobbying, all of which can influence budget line items. We gather information on the DOE's funding requests and Congressional appropriations at the technology level each year to construct these windfalls, and use them to form an instrument for observed funding. This allows us to avoid the restrictive assumptions typically required in the production function estimation literature. In ongoing work, we are extending this research design to firm-level data from Compustat to study how funding

ultimately affects the innovation decisions and economic performance of private-sector firms.

We operationalize this research design using a recentered instrument following Borusyak and Hull (2023). First, to isolate idiosyncratic shocks, we predict appropriations using a flexible function of observables: the DOE request, political control of Congress and the Presidency, and technology characteristics (e.g., whether a technology is "dirty" or politically salient). The residual from this prediction captures the unexpected part of the appropriation and forms the basis of our instrument. To further guard against bias from non-random exposure to funding shocks, we simulate counterfactual appropriation paths by permuting shocks across technologies and years, compute the expected stock of funding under these permutations, and use the deviation from that expectation as our final instrument. This approach strengthens causal interpretation by removing endogenous variation linked to the scale or salience of particular technologies. This empirical strategy enables a unified analysis of how funding shapes the training of scientists, the generation of patentable ideas, and the subsequent activity of firms.

Our paper contributes to the literature on skilled labor supply. A growing body of work documents how enhancing human capital increases and directs innovation (e.g., Bianchi and Giorcelli 2020), as well as how immigrants increase innovation (e.g., Bernstein, Diamond, McQuade, et al. 2019). Our paper also complements recent work examining how scientists choose topics to pursue and how the costs of switching can be high (e.g., Azoulay, Fons-Rosen, and Graff Zivin 2019; Deming and Noray 2020; Myers 2020). Babina, He, Howell, et al. (2023) study how large cuts to federal funding affect university researchers and the innovation they produce. This literature so far has focused on how existing scientists and inventors respond to incentives. By using new data on PhD dissertations, we are able to study the initial choice to become a scientist or inventor.

Our focus on energy technologies builds on literature from energy and environmental economics. There is a long literature on induced innovation showing that demand-side mechanisms like carbon pricing induce clean energy innovation in aggregate and at the firm level (Aghion, Dechezleprêtre, Hémous, et al. 2016; Calel 2020; Calel and Dechezleprêtre 2016; Johnstone, Haščič, and Popp 2010; Newell, Jaffe, and Stavins 1999; Popp 2002, 2019).[2] However, there is less research on investment in human capital itself and the supply of inventors.

2. Popp (2019) provides a comprehensive review of the last decade of research on induced energy and environmental innovation.

Finally, our ongoing work links funding shocks to firm-level outcomes using Compustat, allowing us to examine how government investments in R&D diffuse into the private sector. Our empirical approach builds on Bloom, Schankerman, and Van Reenen (2013), extending it by incorporating the effect of government funding on firm-level decisions and outcomes. It also complements recent work by Fieldhouse and Mertens (2023), who use new time series methods to estimate the impact of government funding on private-sector TFP growth, and Dyevre (2024), who uses firm-level data to study the impact of public and private R&D on long-run productivity growth.

2 Data

2.1 DOE Funding

Our empirical strategy relies on the amount of funding requested by the DOE and the amount that is appropriated by Congress. We gather both pieces of information at the technology-year level from annual statistical tables from the DOE that report requested and actual investment levels. We standardize technology funding categories over time—which tend to change year-to-year—building a new data set of technology-specific funding data from 1980 to 2020.

2.2 Dissertations

We use data from the ProQuest Dissertations and Theses Global Database to construct a dataset of PhD dissertations produced by U.S. higher education institutions. The ProQuest database contains dissertations from all disciplines across 2,315 institutions in 96 countries. We focus on those granted in the U.S. between 1980 and 2020 to align with our DOE funding and Congressional appropriations data. In total, our sample contains dissertations from 809 US universities, with an average of 96 dissertations per institution granted each year.

To classify dissertations into disciplines, we search for the names of disciplines through the text available in the Proquest fields "subject terms," "paper keywords," and "paper categories." We also categorize them as being in STEM disciplines versus the social sciences as shown in Appendix Table A1. To identify dissertations studying energy technology-related topics, we construct another dictionary of terms specific to different energy types, such as "solar panel" and "photovoltaic" for

solar energy.[3] We then search for these terms in each dissertation's title, abstract, and keywords.[4]

We follow prior literature to classify different energy types into clean, dirty, and grey (i.e., those that could improve the environmental performance of a technology that is otherwise dirty) as follows:[5]

- **Clean:** Renewables (Solar, Wind, Geothermal, Hydro, Marine), Carbon Capture, Nuclear, Smart Grid, Energy Storage, Battery, Fuel Cell, Electric Vehicles, Hybrid Vehicles
- **Dirty:** Fossil Fuels, Oil, Natural Gas Coal, Internal Combustion Engine
- **Grey:** Energy Efficiency, Biomass and Waste

Since some dissertations have keywords that cross technology and category boundaries, we count the number of occurrences of each term or combination of terms in all relevant fields and use this information to classify dissertations by individual technologies as well as broader classifications of clean, dirty, or grey. If a dissertation contains terms related to, say, both solar and wind energy, we count this dissertation both as solar and wind, using weighted counts proportional to the share of solar and wind terms.

2.3 Patents

We gather patent data from PATSTAT and identify energy technology-specific patents using an approach that is similar to our approach for dissertations.

2.4 Firm Inputs and Outputs

We collect data from Compustat on the inputs and outputs of publicly-listed private sector firms. We link individual firms to energy technologies in two steps. First, we utilize the firm-to-patent mapping developed by Dyevre and Seager (2023). This mapping links publicly-listed firms in Compustat to patents from the U.S. Patent and Trademark Office (USPTO) over the period 1950 to

3. Appendix Table A2 provides a complete list of our energy terms. We standardize strings by making all letters lowercase and replacing special characters (e.g., "-") with a space before searching for keywords.

4. The keywords are a list of terms provided by the dissertation's author.

5. Our classification aligns with others developed in the literature so far, such as in Popp, Pless, Haščič, et al. (2020) and Aghion, Dechezleprêtre, Hémous, et al. (2016).

2020. Second, we use the patent-to-technology we developed for this project to link firms' patents to technologies. In combination, these two linkages allow us to assign firms to the technologies based on their patent portfolios. This allows us to relate our data on DOE funding levels and shocks at the technology level to the firms that are most relevant based on their similarity in technology space.

3 Trends in DOE Funding, Dissertations, and Patenting

Before assessing how DOE funding windfalls impact dissertations and patents, we start by exploring the data descriptively and document trends that emerge. In Figure 1, we plot DOE R&D obligations broken down into three recipient categories: government and non-profit organizations, private businesses, and higher education. Roughly 40 to 60 percent of R&D obligations went towards higher education from 2000 through 2018, suggesting that DOE expects research conducted by those at universities to play an important role in driving energy innovation.

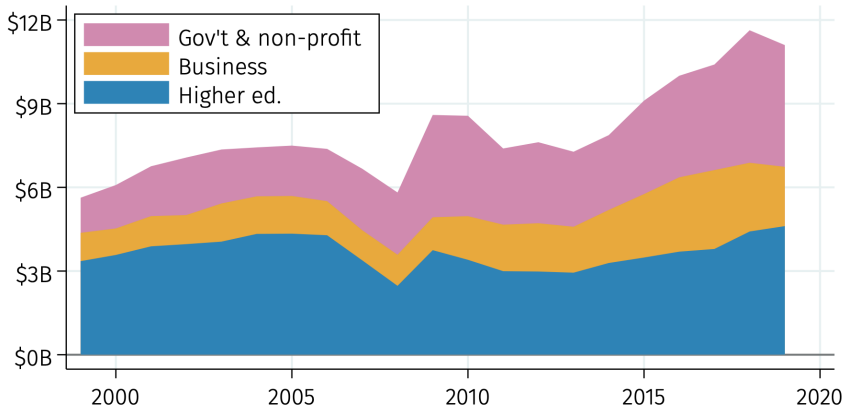


Figure 1
Department of Energy R&D Obligations by Organization Type

The main source of variation that we exploit in our research design are funding windfalls—the difference between the amount of funding Congress appropriates and what DOE requested—at the technology-year level. Figure 2 illustrates this variation for five groups of technologies that we study. There is significant variation both over time and across technologies.

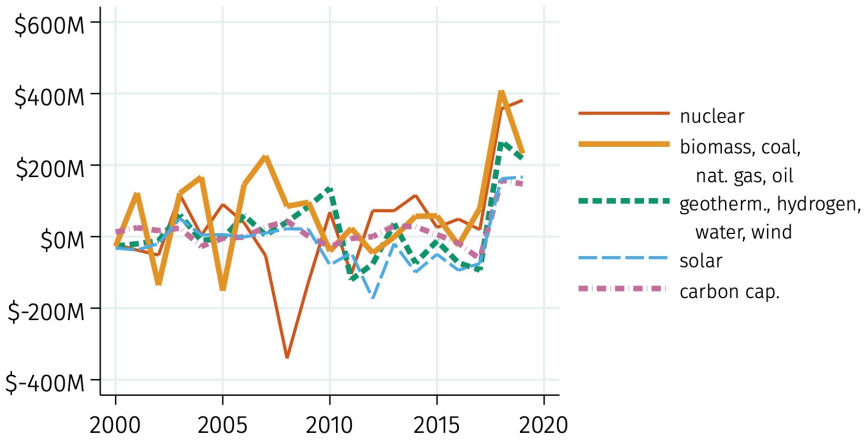


Figure 2
Technology-Specific Funding Windfalls (Appropriations Minus Requests)

In Figure 3, we plot the share of all dissertations that are related to clean, dirty, and grey energy topics, and we plot the counts of new dissertations in each category in Figure 4. These figures illustrate two key points. First, the share of researchers focusing on clean energy has been steadily increasing since about the year 2000. This is driven by an increase in the number of researchers, not just a change in the composition.

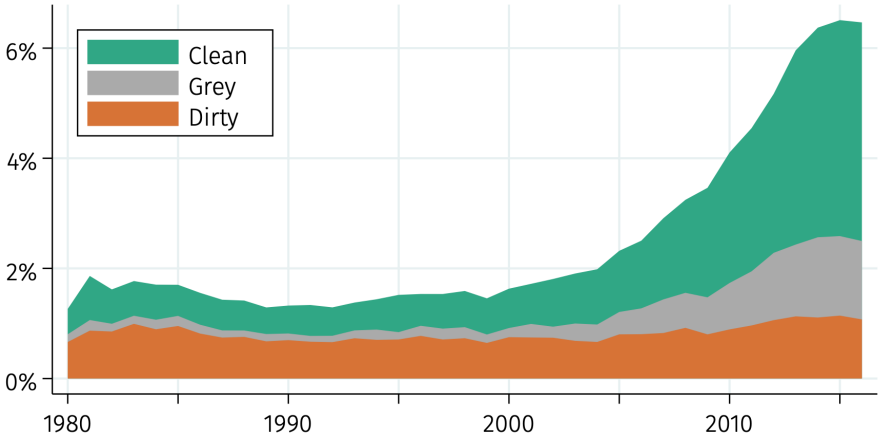


Figure 3
Energy-Focused Dissertations Over Time as Share of all PhD Dissertations

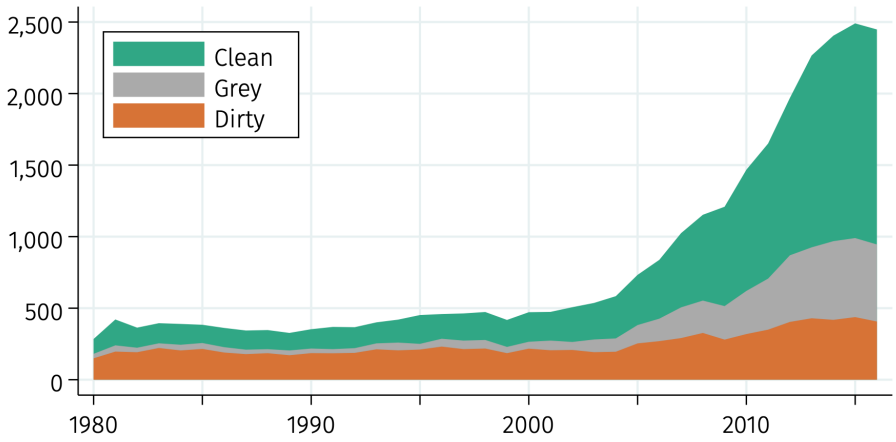


Figure 4
Number of New Energy-Focused Dissertations Over Time

In Figure 5, we plot total DOE R&D funding (in millions of dollars), dissertation counts, and patent counts for clean versus dirty energy. These patterns could, of course, be driven by many factors. Our objective in our empirical analysis is to isolate the causal effect of DOE funding on the creation of new scientists.

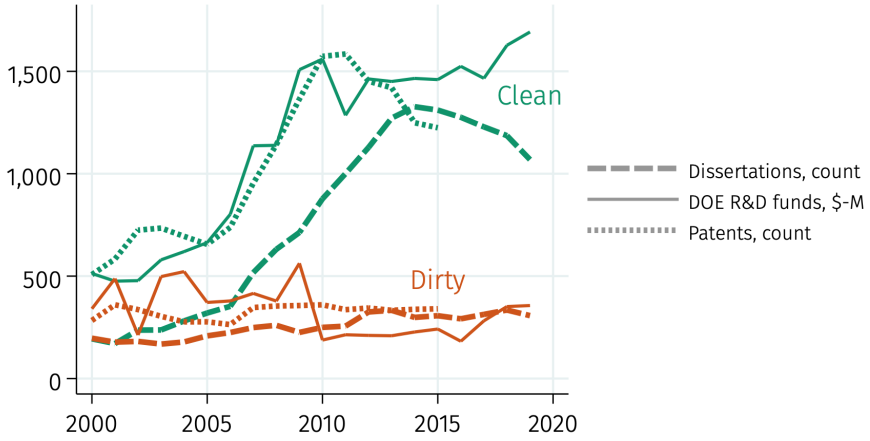


Figure 5
Dissertations, Patents, and DOE R&D Funding

4 Empirical Strategy Overview

We use a production function framework to model how DOE funding affects the quantity of scientists (i.e., dissertations) and science (i.e., patents) that are focused on specific energy technologies each year. The production function relates the flow of outcomes (Y) in a given technology (j) and year (t) to the stock of DOE R&D investments (K), which is inspired by the knowledge production function of Griliches (1979). As in the prior literature, we use lower-case variables to indicate logged values. The log-transformed production function is:

$$y_{jt} = \beta_0 + \beta_1 k_{jt} + \omega_{jt} + \varepsilon_{jt}\,, \tag{1}$$

where ω_{jt} captures technology-year-specific productivity shocks and ε_{jt} are idiosyncratic shocks. The presence of unobserved (to the econometrician) determinants of output, ω_{jt}, may bias our estimates of β_1 if not accounted for.

A long methodological literature on production function estimation has emerged to address this simultaneity bias (e.g., Ackerberg, Caves, and Frazer 2015; Levinsohn and Petrin 2003; Olley and Pakes 1996).[6] There are two important differences between our setting and more conventional settings studied in that prior literature. First, we only have one input. This simplifies our estimation strategy and circumvents the challenge of separately identifying the elasticities of output with respect to multiple inputs.

Second, and more importantly, our setting has the unique feature that realized investments differ from the amount requested by DOE staff due to the nature of the Congressional appropriations process. The amount of funding appropriated by Congress differs from the amount that is requested, as final budgeting decisions emerge from a negotiating process that involves hearings to question officials about their requests followed by budget resolutions drafted by Congressional committees, which then must be passed by both the House and Senate and thus typically also entails another set of negotiations between the House and Senate. Final technology-specific funding amounts are therefore affected by quasi-experimental variation from the political process, such as legislators' personal and local preferences, as well as budget shocks elsewhere. The process frequently takes anywhere from six months to a year or longer. We use Congressional funding windfalls as an

6. Production function estimation is also subject to selection bias, but that does not apply to our setting.

instrument for the stock of DOE funding.

Implementing the windfall instrument. We operationalize Congressional windfalls by first predicting technology–year appropriations A_{jt} from the DOE request and political variables that summarize the budget process. Specifically, we estimate

$$\log A_{jt} \;=\; f(\log R_{jt}, \mathrm{Pol}_t, \mathrm{Dirty}_j) \;+\; \varepsilon_{jt}, \tag{2}$$

where R_{jt} is the DOE request; Pol_t includes indicators for party control of the Presidency, Senate, and House; and Dirty_j flags fossil-fuel technologies. The function $f(\cdot)$ is specified flexibly to capture how appropriations co-move with requests and political conditions, and to allow for heterogeneous political effects across technologies and over time. In practice, we use a linear function with: (i) interactions of $\log R_{jt}$ with year fixed effects (request–year flexibility), (ii) technology and year fixed effects, (iii) interactions between political variables and Dirty_j, and (iv) limited nonlinear terms in $\log R_{jt}$ (such as quadratic and cubic terms). This yields fitted values $\widehat{\log A}_{jt}$ and residuals $\hat{e}_{jt} = \log A_{jt} - \widehat{\log A}_{jt}$.

We construct a "windfall-stock" instrument at the level of accumulated, discounted appropriations. Let the discounted stock of appropriations be:

$$S_{jt} \equiv \sum_{\tau \geq 0} (1-\delta)^{\tau} A_{j,t-\tau-\tau_0}, \tag{3}$$

where δ is the annual depreciation rate that governs how quickly past appropriations lose relevance for current outcomes, and $\tau_0 \in \{0,1\}$ controls the timing alignment between funding and outcomes: when $\tau_0 = 0$ we link outcomes in year t to funding flows realized in the same year, and when $\tau_0 = 1$ we instead link outcomes in year t to funding flows realized in year $t-1$.

We construct a counterfactual predicted stock $\widehat{S}_{jt}$ by replacing A_{jt} with $\widehat{A}_{jt} = \exp(\widehat{\log A}_{jt})$ in the same recursion. The resulting instrument is

$$Z_{jt} \;=\; \log S_{jt} \;-\; \log \widehat{S}_{jt}, \tag{4}$$

which isolates the unexpected component of the funding stock relative to what is predictable based

on requests and political conditions.

Addressing Non-Random Exposure: Recentered Instrument. Equal-percentage shocks can imply much larger absolute "treatments" for technologies with large baseline budgets. To address this non-random exposure, we construct a recentered instrument following Borusyak and Hull (2023). Concretely, we (i) randomly permute the residual shocks $\hat{e}_{jt}$ across technology–years and add them to the fitted values $\widehat{\log A}_{jt}$ to obtain artificial appropriations $\tilde{A}_{jt}^{(n)} = \exp(\widehat{\log A}_{jt} + \tilde{e}_{jt}^{(n)})$; (ii) accumulate each artificial series into a simulated stock $\tilde{S}_{jt}^{(n)}$ using the same (δ, τ_0); and (iii) average over permutations to obtain the expected log stock

$$\mu_{jt} \;=\; \frac{1}{N}\sum_{n} \log \tilde{S}_{jt}^{(n)}.$$

The *recentered instrument* is then

$$z_{jt} \;=\; \log S_{jt} \;-\; \mu_{jt}, \tag{5}$$

which captures quasi-random bargaining noise net of systematic exposure to shock magnitudes.

Estimation and Interpretation. The second stage estimates

$$\log Y_{jt} \;=\; \beta \log S_{jt} \;+\; X_{jt}'\theta \;+\; \varepsilon_{jt}, \tag{6}$$

instrumenting $\log S_{jt}$ with either Z_{jt} or z_{jt}. Outcomes Y_{jt} correspond to the flows of PhD dissertations or patent applications. Controls X_{jt} include year fixed effects, technology fixed effects, and political variables interacted with Dirty_j. We report results across $\delta \in \{0, 0.1, 0.2, 0.5, 0.7\}$ and $\tau_0 \in \{0, 1\}$, and provide complementary distributed-lag specifications as diagnostics for dynamic responses and to guide the choice of δ. Identification rests on the assumption that, conditional on requests and observed politics, congressional deviations reflect bargaining shocks orthogonal to unobserved productivity ω_{jt}, and that Congress does not systematically possess superior technology-specific information relative to the DOE. For estimates based on the recentered instrument, we report permutation-based standard errors aligned with the instrument's construction.

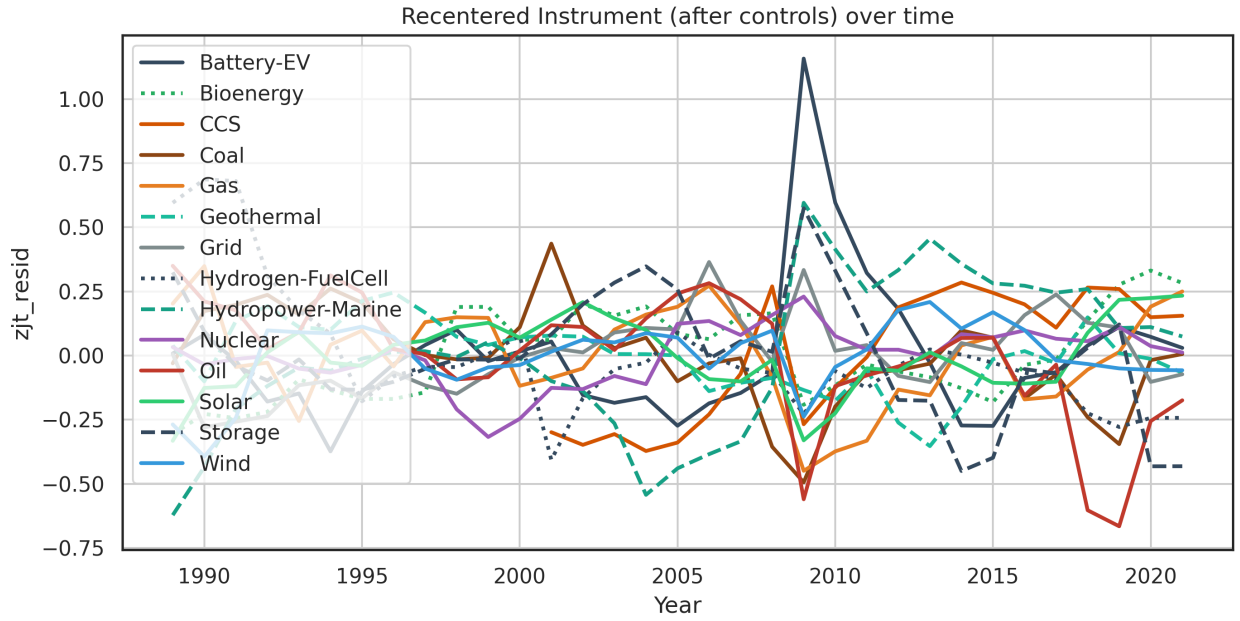


Figure 6
Recentered instrument z_{jt} by technology, benchmark specification with discount rate $\delta = 0.5$ and contemporaneous timing ($\tau_0 = 0$). Values are residualized on year and technology fixed effects and political controls. The pronounced spike around 2009 reflects the ARRA-related funding surge.

4.1 Preliminary Results

Prediction step. Appropriations co-move tightly with requests and observed political conditions. Across increasingly flexible specifications, the prediction of $\log A_{jt}$ attains R^2 in the 0.92–0.96 range, indicating that our residual shocks isolate a relatively small, unexpected component of appropriations.

Instrument construction and recentering. Using the residual-based stock instrument $Z_{jt} = \log S_{jt} - \log \widehat{S}_{jt}$, we then apply the Borusyak–Hull recentering procedure to address non-random exposure. For our benchmark specification with a discount rate $\delta = 0.5$ and contemporaneous alignment ($\tau_0 = 0$), Figure 6 plots the recentered instrument $z_{jt} = \log S_{jt} - \mu_{jt}$ after netting out the full set of controls (year and technology fixed effects and political variables). The figure shows substantial variation across technologies and a pronounced spike around 2009, which reflects the large funding surge associated with the American Recovery and Reinvestment Act (ARRA). Alternative discount rates δ redistribute the identifying variation across periods in intuitive ways by changing how strongly the instrument weights past appropriations.

Main IV results. Two-stage least squares estimates using z_{jt} indicate positive elasticities for both dissertation flows and patenting with respect to the stock of DOE appropriations. Across specifications that include year and technology fixed effects as well as political controls, the estimated coefficients cluster in the 0.2–0.4 range (see Figures 7–8). The IV estimates are close to the corresponding OLS coefficients, suggesting that endogeneity in funding is not generating large upward or downward biases in simple correlations. When we vary the depreciation rate $\delta \in \{0.1, 0.2, 0.5, 0.7\}$ and the timing parameter $\tau_0 \in \{0, 1\}$, we obtain a consistent qualitative pattern. Lower discount rates place greater weight on older appropriations and tend to yield somewhat larger but less precisely estimated coefficients.

Back-of-the-envelope magnitudes (preliminary). Under our benchmark specification, a one-time increase in the appropriations *stock* of about $250 million is associated with roughly one additional patent per year and one additional PhD per year thereafter. This calculation treats the stock shift as persistent; if instead it is generated by a single-year flow, the implied annual effects would decline at the discount rate δ. These magnitudes should be interpreted cautiously: our appropriations measure aggregates all DOE R&D spending (including national laboratories and other mission activities), whereas our outcomes focus on university-centered outputs (PhD dissertations) and USPTO patenting. Because not all appropriated dollars directly target these specific outcomes, the implied effect per appropriations dollar is attenuated relative to directed funding.

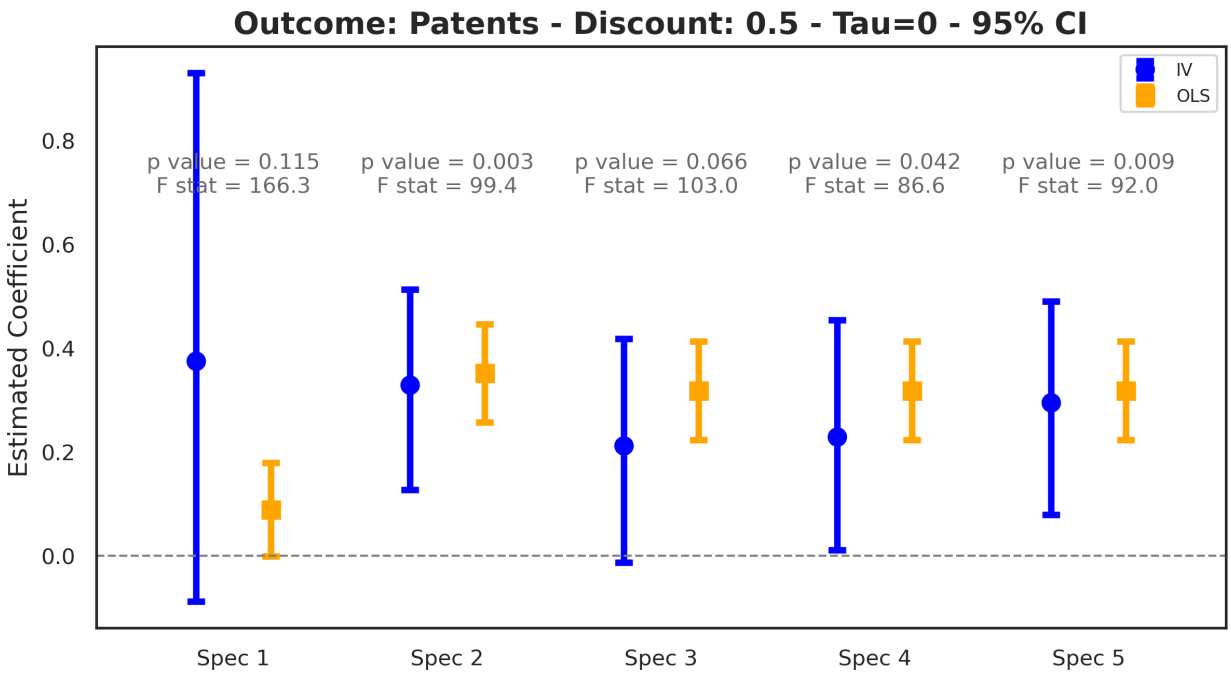


Figure 7
IV coefficients across specifications ($\delta = 0.5$, $\tau_0 = 0$).

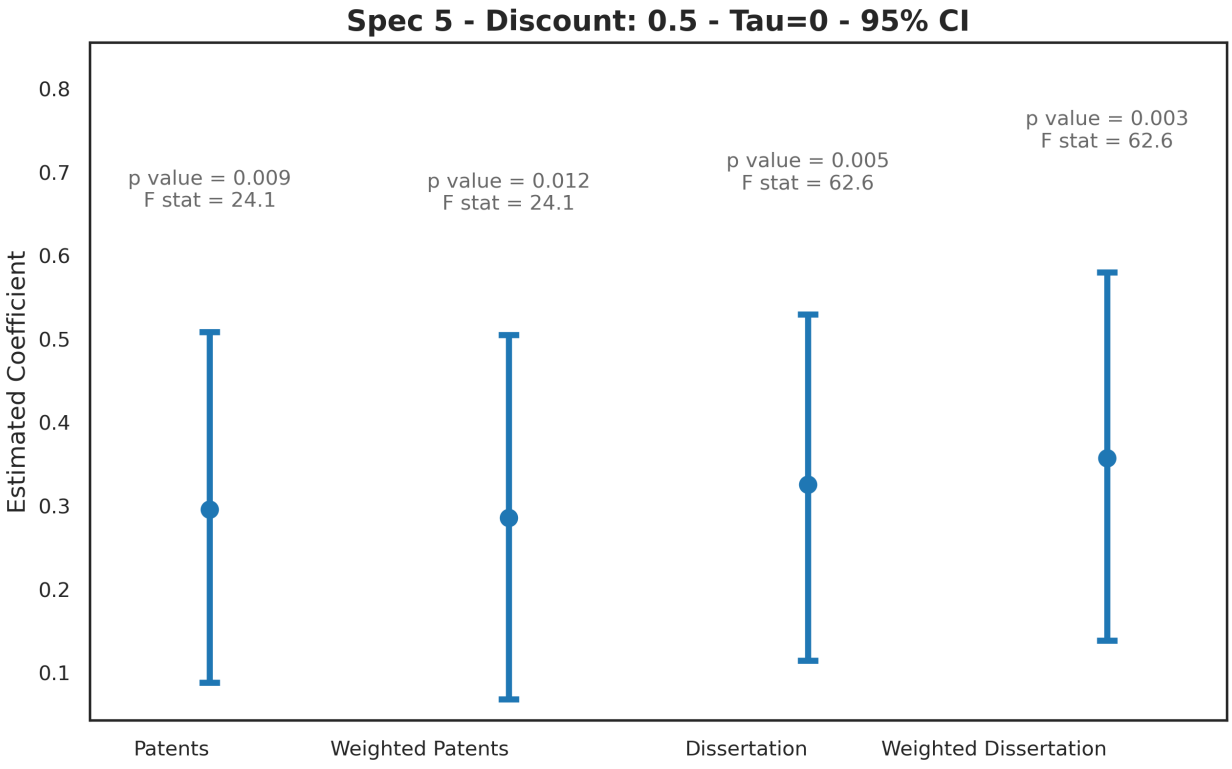


Figure 8
IV coefficients across outcomes (dissertations, patents).

5 Ongoing Work

We propose extending our analysis of Department of Energy (DOE) funding windfalls to examine firm-level economic outcomes, specifically studying the extent to which public R&D funding acts as a complement to, or a substitute for, private R&D investment. While our current results demonstrate that DOE appropriations causally increase the supply of specialized human capital (energy-related dissertations) and scientific outputs (patents), the transmission of these shocks to the broader economy remains unexamined. Existing macro-level evidence suggests positive returns to government R&D (Fieldhouse and Mertens 2023), yet the micro-foundations of this growth—specifically how individual firms adjust their R&D budgets in response to public funding shocks—are not well understood. We will analyze two mechanisms underlying these aggregate impacts: first, how public funding affects productivity directly; and second, how public funding affects productivity through firms' decisions to adjust their R&D budgets in response to public funding shocks. The second channel could either amplify or dampen the first depending on the extent to which public R&D spending crowds in or crowds out private R&D spending.

To test this, we will integrate our dataset of exogenous DOE funding windfalls with firm-level data from Compustat described in section 2. Building on the framework of Bloom, Schankerman, and Van Reenen (2013), we will introduce the stock of public R&D funding into a firm's R&D augmented production function. We will analyze its direct effect on productivity as well as its indirect effect through firms' R&D factor demands. To do so, we will construct a shift-share instrument that interacts a firm's pre-determined technology mix (inferred from their patent portfolio) with our technology-specific DOE funding windfalls. This approach will allow us to isolate the causal impact of public funding shocks on private R&D spending and productivity, distinct from endogenous demand shocks.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–2451. JSTOR: 43866416.

Aghion, Philippe, Antoine Dechezleprêtre, David Hémous, Ralf Martin, and John Van Reenen. 2016. "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry." *Journal of Political Economy* 124 (1): 1–51.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60 (2): 323–351. JSTOR: 2951599.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *American Economic Review* 109 (8): 2889–2920.

Babina, Tania, Alex Xi He, Sabrina T Howell, Elisabeth Ruth Perlman, and Joseph Staudt. 2023. "Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneurship, and Publications*." *The Quarterly Journal of Economics* 138 (2): 895–954.

Bernstein, Shai, Rebecca Diamond, Timothy McQuade, and Beatriz Pousada. 2019. *The Contribution of High-Skilled Immigrants to Innovation in the United States.* Working Paper.

Bianchi, Nicola, and Michela Giorcelli. 2020. "Scientific Education and Innovation: From Technical Diplomas to University Stem Degrees." *Journal of the European Economic Association* 18 (5): 2608–2646.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica* 81 (4): 1347–1393.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A Toolkit of Policies to Promote Innovation." *Journal of Economic Perspectives* 33 (3): 163–184.

Borusyak, Kirill, and Peter Hull. 2023. "Nonrandom Exposure to Exogenous Shocks." *Econometrica* 91 (6): 2155–2185.

Calel, Raphael. 2020. "Adopt or Innovate: Understanding Technological Responses to Cap-and-Trade." *American Economic Journal: Economic Policy* 12 (3): 170–201.

Calel, Raphael, and Antoine Dechezleprêtre. 2016. "Environmental Policy and Directed Technological Change: Evidence from the European Carbon Market." *The Review of Economics and Statistics* 98 (1): 173–191.

Deming, David J., and Kadeem Noray. 2020. "Earnings Dynamics, Changing Job Skills, and STEM Careers." *The Quarterly Journal of Economics* 135 (4): 1965–2005.

Dyevre, Arnaud. 2024. *Public R&D Spillovers and Productivity Growth.* Working Paper. LSE.

Dyevre, Arnaud, and Oliver Seager. 2023. *70 Years of Patents Matched to Firms: Methodology and Insights About Firm Heterogeneity.* Working Paper.

Fieldhouse, Andrew, and Karel Mertens. 2023. *The Returns to Government R&D: Evidence from U.S. Appropriations Shocks.* SSRN Scholarly Paper, 4452881, Rochester, NY. Social Science Research Network: 4452881.

Goolsbee, Austan. 1998. "Does Government R&D Policy Mainly Benefit Scientists and Engineers?" *The American Economic Review* 88 (2): 298–302. JSTOR: 116937.

Griliches, Zvi. 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." *The Bell Journal of Economics* 10 (1): 92–116. JSTOR: 3003321.

Johnstone, Nick, Ivan Haščič, and David Popp. 2010. "Renewable Energy Policies and Technological Innovation: Evidence Based on Patent Counts." *Environmental and Resource Economics* 45 (1): 133–155.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70 (2): 317–341. JSTOR: 3648636.

Myers, Kyle. 2020. "The Elasticity of Science." *American Economic Journal: Applied Economics* 12 (4): 103–134.

Newell, Richard G., Adam B. Jaffe, and Robert N. Stavins. 1999. "The Induced Innovation Hypothesis and Energy-Saving Technological Change." *The Quarterly Journal of Economics* 114 (3): 941–975.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263–1297.

Popp, David. 2002. “Induced Innovation and Energy Prices.” *The American Economic Review* 92 (1): 160–180. JSTOR: 3083326.

———. 2019. *Environmental Policy and Innovation: A Decade of Research.* Working Paper 25631. National Bureau of Economic Research.

Popp, David, Jacquelyn Pless, Ivan Haščič, and Nick Johnstone. 2020. *Innovation and Entrepreneurship in the Energy Sector.* Working Paper, Working Paper Series 27145. National Bureau of Economic Research.

Romer, Paul M. 1990. “Endogenous Technological Change.” *Journal of Political Economy* 98 (5): S71–S102. JSTOR: 2937632.

———. 2000. “Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers?” *Innovation policy and the economy* 1:221–252.

Van Reenen, John. 2021. *Innovation and Human Capital Policy.* Technical report w28713. National Bureau of Economic Research.

A Appendix

Table A1
Disciplines List

Social Sciences
Archaeology
Area planning and development
Commerce-Business
Criminology
Cultural anthropology
Demography
Economic history
Economic theory
Economics
Folklore
Forensic anthropology
Geography
Individual & family studies
International relations
Labor economics
Labor relations
Military studies
Organization theory
Organizational behavior
Peace studies
Physical anthropology
Political Science
Public administration
Public policy
Recreation
Social research
Social structure
Social work
Sociolinguistics
Sociology
Transportation
Urban planning

STEM
Agriculture
Agriculture economics
Agriculture education
Agriculture engineering
Agronomy
Animal diseases
Animal sciences
Aquatic sciences
Forestry
Horticulture
Plant pathology
Plant sciences
Range management
Soil sciences
Urban forestry
Wildlife management
Biochemistry
Bioinformatics
Biology
Biomechanics
Biophysics
Biostatistics
Botany
Cellular biology
Developmental biology
Endocrinology
Entomology
Evolution & development
Genetics
Histology
Limnology
Microbiology
Molecular biology
Morphology
Neurosciences
Parasitology
Physiology
Systematic biology
Virology
Zoology
Ecology
Macroecology
Paleoecology
Aerospace engineering
Artificial intelligence
Automotive engineering
Biomedical engineering
Bioengineering
Chemical engineering
Civil engineering
Computer engineering
Computer science
Electrical engineering
Engineering
Geological engineering
Geophysical engineering
Geotechnology
Hydraulic engineering
Industrial engineering
Mechanical engineering
Mining
Naval engineering
Nanotechnology
Nuclear engineering
Ocean engineering
Operations research
Packaging
Petroleum engineering
Plastics
Robotics
System science
Conservation biology
Environmental economics
Environmental education
Environmental engineering
Environmental geology
Environmental health
Environmental justice
Environmental law
Environmental management
Environmental philosophy
Environmental science
Environmental studies
Land use planning
Natural resource management
Water resources management
Wildlife conservation
Aeronomy
Atmospheric chemistry
Atmospheric sciences
Biogeochemistry
Biological oceanography
Chemical oceanography
Continental dynamics
Geobiology
Geochemistry
Geographic information science and geodesy
Geology
Geomorphology
Geophysics
Hydrologic sciences
Marine geology
Meteorology
Mineralogy
Paleoclimate science
Paleontology
Petroleum geology
Petrology
Physical geography
Physical oceanography
Planetology
Plate tectonics
Remote sensing
Sedimentary geology
Aging
Alternative medicine
Audiology
Dentistry
Epidemiology
Gerontology
Health care management
Health sciences
Immunology
Kinesiology
Medical ethics
Medical imaging
Medicine
Mental health
Nursing
Nutrition
Obstetrics
Occupational safety
Occupational therapy
Oncology
Ophthalmology
Optometry
Osteopathic medicine
Pathology
Pharmaceutical sciences
Pharmacology
Physical therapy
Public health
Public health education
Speech therapy
Surgery
Therapy
Toxicology
Veterinary medicine
Acoustics
Analytical chemistry
Applied mathematics
Applied physics
Astronomy
Astrophysics
Atomic physics
Chemistry
Computational chemistry
Computational physics
Condensed matter physics
Electromagnetics
Fluid mechanics
High energy physics
High temperature physics
Inorganic chemistry
Low temperature physics
Materials science
Mathematics
Mechanics
Molecular chemistry
Molecular physics
Nanoscience
Nuclear chemistry
Nuclear physics and radiation
Optics
Organic chemistry
Physical chemistry
Physics
Plasma physics
Polymer chemistry
Quantum physics
Statistics
Statistical physics
Thermodynamics
Theoretical mathematics
Theoretical physics

Table A2
Energy Term Dictionary

Solar
solar energy
solar power
photovoltaic
solar thermal energy
solar tower
solar cell
solar plant
solar panel

Wind
wind turbine
wind farm
wind energy
wind power

Geothermal
geothermal

Hydro
hydropower
hydro power
hydro energy
hydroelectric
hydro electric
hydro plant

Marine
wave energy
wave power
tidal energy
tidal power
tidal turbine
ocean thermal energy
salinity gradient power
salinity gradient energy

Other Renewables
renewable energy
alternative energy

Carbon Capture
carbon capture
carbon sequestration
carbon storage
carbon dioxide capture
carbon dioxide sequestration
carbon dioxide storage
co2 capture
co2 sequestration
co2 storage
carbon dioxide + disposal

Nuclear
nuclear power
nuclear reactor
nuclear energy
nuclear fuel
nuclear engineering
nuclear waste
nuclear accident
nuclear safety
nuclear fission
radioactive waste
nuclear reprocessing

Smart Grid
smart grid
smart power grid
power grid
microgrid
distributed energy

Battery
battery pack
battery module
battery cell
battery lithium ion
battery li ion
battery lead acid
battery electrode
battery storage
battery electrochemical
battery cathode
battery anode
battery discharge
battery rechargability
lead acid + battery
li-ion + battery
lithium ion + battery
li ion + battery
lithium-ion + battery
pack + battery
module + battery
cell + battery
electrode + battery
storage + battery
electrochemical + battery
cathode + battery
anode + battery
discharge + battery
rechargeability + battery

Fuel Cell
fuel cell

Energy Storage
energy storage
flywheel
supercapacitor

Electric Vehicles
electric vehicle
electric mobility
electrical propulsion units
electric propelled vehicle
electrodynamic brake
electric regenerative braking
charging station

Hybrid Vehicles
hybrid vehicle
hybrid propulsion system
regenerative braking

Biomass And Waste
biomass energy
biorefinery
lignocellulosic biomass
bioenergy
biofuel
biodiesel
bio waste
syngas
biochar
bioethanol
waste heat
waste energy
waste incineration

Energy Efficiency
energy efficiency
energy efficient
energy conservation
energy saving
energy recuperation

Fossil Fuels
fossil fuel
combustion furnace
liquid fuel
carbonaceous
steam engine
steam accumulator
steam generation
kiln + engineering

Internal Combustion Engine
combustion engine
combustion + energy
gasoline + engine

Oil
crude oil
gasoline
diesel
oil refinery
jet fuel

Natural Gas
natural gas
gaseous fuel
synthetic natural gas
liquefied pretroleum gas
gas turbine

Coal
solid fuel
coal

Note: When the keyword is a combination of terms (e.g., "lead acid + battery"), we search for joint occurrences, that is, a dissertation would be classified as Battery if it contains the phrase "lead acid" and "battery" even if there are not immediately next to each other.