Causes and Consequences of Innovation Clusters: Examining the Impact of National Labs*

Susan Helper, Resem Makan, and Daniel Shoag†

Case Western Reserve University

3 October 2024

Abstract

We report some early results from an NSF-funded project on understanding the origins and outcomes of technology and science-based regional ecosystems. In particular, we look at the impacts of large federal investments in 17 National Laboratories on both innovation- and equity-related outcomes in surrounding geographies. Using synthetic controls and "runner-up" locations, we find that the establishment of a lab in most cases leads to significantly more patenting in the host county, compared to similar regions, and that the differences grew over time. We find suggestive evidence of sustained long-run increases in wages, compared to runner-up counties. These results are consistent with a hypothesis that labs had place-based spillover effects (but not conclusive). In future research we will do much more to explore alternative explanations. For example, we will link lab openings in the 1940s to declassified individual Census records to look at labor market flows. In several cases we can combine this data with archival data, to track precise residential locations of lab employees. Finally, we intend to use digitized City Book data to track the impact of the labs on business formation.

Keywords: Innovation Clusters, National Laboratories, R&D, Government Investments, Placed-based policies, Knowledge Spillovers.

*This is a preliminary draft and is not meant to be circulated.

†We gratefully acknowledge the research assistance provided by Dawson Fishman, Megan Abel, and John McCormick. We also thank the National Science Foundation (NSF) for funding this project.

1 Introduction

Last month, we received an NSF grant to study innovative eco-systems. In this paper we present some early results that are quite robust and are intriguing (at least to us). If chosen to present in December, we would have a proper paper.

A variety of academic and policy proposals seek to make significant investments to catalyze technology and science-based regional ecosystems that will provide broad economic and societal benefits. For example, Gruber and Johnson (2019) propose to "jump-start America" by creating a competition for municipalities to win federal grants to become new technology hubs. The U.S. Congress has appropriated billions of dollars for programs such as Tech Hubs (heavily influenced by the Gruber and Johnson proposal), RECOMPETE, and the Build Back Better Regional Challenge (and has authorized the spending of billions more, in the Chips and Science Act). For example, the NSF Engines program "aims to fund regional coalitions of partnering organizations to catalyze technology and science-based regional innovation ecosystems" (NSF Dear Colleague Letter).[1] The European Union and China have also invested in innovative eco-systems (Murphree and Breznitz, 2020; Mazzucato, 2021).

While there is significant research on economic development and public investments, this work has been siloed by a) methodology (quantitative/historical vs quantitative), and b) outcomes of interest (innovation vs. poverty alleviation).[2] Further, the existing literature generally does not study public investments comparable in size to those currently underway or being contemplated.

Our approach will attempt to bridge these siloes. We will use nuanced qualitative analysis of case studies to inform a hybrid narrative-statistical approach based on counterfactuals and to consider metrics related to both innovation and broad-based prosperity. Ultimately, we aim to build a broad based, statistical model that can be applied for forecasting.

To understand the impacts of efforts to create innovative eco-systems, we would ideally like to find historical analogues to such programs. In this paper we examine the impact of national laboratories . While the initial purpose of these labs was generally not to build local innovation, we argue that they had many other similarities to current cluster-based programs and proposals: the investments were aimed at promoting innovation; were concentrated in local geographies; and often were quite large relative to the local economies in which they were established. Several features of the program allow us to tease out causal implications of these programs for innovation and broad-based prosperity:

- In many cases, we can identify runner-up locations, allowing us to compare trends

[1] see NSF Dear Colleague Letter NSF, 24-036.

[2] See, for example, (Saxenian, 1994; Busso et al., 2013; Greenstone et al., 2010) and work in progress by Slattery (2022).

before and after the establishment of a lab for these locations.

- Dozens of labs were phased out, closed, or decertified. Again, assigning counterfactual locations, the same exercise could be constructed in reverse.

- The initial and on-going budgets of the programs varied significantly (allowing understanding of a dose-response function).

A complicating factor is that national labs are rarely the only economic-development related "treatment" that a region undergoes. We will use our counterfactual locations and qualitative analysis to understand the relationship between the establishment of a lab in a region and subsequent additional investments, asking questions such as: did the existence of a lab "crowd in" further investment in a region? Were lab investments and other investments responding to some common factor that our analysis did not control for? These questions are especially difficult when labs are established in areas that already have some level of R&D investments.

We address these issues by using counterfactuals constructed from a) runner-up locations and b) synthetic controls. Our results so far offer suggestive support for the idea that government investments can in fact "jump start" innovation ecosystems. Counties with labs patent significantly more than counterfactual counties; the differences are larger when lab investment is greater, and grow over time, including patenting by non-lab employees. We find suggestive results that the establishment of a lab leads to persistently higher wage growth in host counties compared to counterfactual counties.

Below we discuss previous research in this area, provide background on the national laboratories, describe our data and methods, and present our results. These results are highly preliminary; our conclusion discusses next steps we plan to take.

2 Background on National Labs

Most studies of the impact of research institutions focus on universities. In this paper, we focus on the economic impact of U.S. national laboratories. This focus is useful for several reasons. The formation of a university involves a complicated mix of treatments over decades and even centuries. The national laboratory program, which is focused solely on research, is perhaps a more accurate approximation of government programs that focus on applied science. While the laboratories do host post-docs, they do not house tens of thousands of students. Unlike universities, they are much more tied to federal funding and a federal agency (the Department of Energy). Understanding where these more targeted institutions

can, nevertheless, have large spillovers is important for understanding how to craft ongoing interventions like the NSF Regional Innovation Engines program.

All but one of the 17 national laboratories are Federally Funded Research and Development Centers (FFRDCs).[3] See Table-1 for a list of the 17 national labs. Today, these are overseen by the U.S. Department of Energy (which was not founded until 1977).

Table 1: List of National Labs and their locations

National Lab	**County, State**
Argonne National Laboratory	DuPage County, Illinois
Brookhaven National Laboratory	Suffolk County, New York
Argonne National Laboratory	DuPage County, Illinois
Fermi National Accelerator Laboratory	Suffolk County, Illinois
Frederick National Laboratory for Cancer Research	Frederick County, Maryland
Idaho National Laboratory	Bonneville County, Idaho
Lawrence Berkeley National Laboratory	Alameda County, California
Lawrence Livermore National Laboratory	Alameda County, California
Los Alamos National Laboratory	Los Alamos County, New Mexico
National Renewable Energy Laboratory	Jefferson County, Colorado
Oak Ridge National Laboratory	Anderson County, Tennessee
Pacific Northwest National Laboratory	Benton County, Washington
Princeton Plasma Physics Laboratory	Middlesex County, NJ
Sandia National Laboratories	Bernalillo County, New Mexico
Savannah River National Laboratory	Aiken County, South Carolina
SLAC National Accelerator Laboratory	Santa Clara County, California
Thomas Jefferson National Accelerator Facility	Newport News, Virginia

The National Labs are owned and funded by the federal government but operated by contractors. The goal is to provide a balance of independence and flexibility that the federal government does not, while still being responsive to the needs of the home agency. The workforce of the National Labs are typically employees of the contractor and not civil servants. A National Lab may be given special access to government and supplier data, employees, and facilities. There is a long-term relationship between a National Lab and its sponsoring agency. Contracts are for 5 years, but may be renewed indefinitely. For example, the Pacific Northwest National Laboratory has been operating under the same contract since 1964 (CRS, 2020; Dale and Moy, 2000).

The National Laboratory system emerged from the extraordinary scientific efforts of World War II, particularly the Manhattan Project, which was tasked with developing the

[3]The exception is the National Energy Technology Laboratory (NETL) which is operated as well as owned by the Department of Energy.

first nuclear weapons. Following the war, the U.S. government maintained these research institutions in an effort to continue advancing scientific knowledge and technology. During the Cold War, national labs played a pivotal role in nuclear research and development, for both nuclear weapons and nuclear energy. It appears that 14 of the 17 national labs were originally created to develop nuclear weapons (or other kinds of advanced weapons).[4] Most came directly out of the Manhattan Project with different labs working on different aspects of that work to produce an atomic bomb – Oak Ridge and Ames producing high concentration uranium, Hanford (now Pacific Northwest National Lab) producing plutonium, Los Alamos developing the actual bomb, and others (Argonne) working on the chemistry and metallurgy required. The labs also studied radiation and how radiation affected humans, instruments for measuring radiation, explosion physics, etc.

After World War II, several more labs were created to develop hydrogen fusion bombs (Livermore, Sandia, Savannah River, etc.), nuclear reactors for submarines (Idaho) and then for navy ships, and eventually for commercial use. After nuclear bomb testing was outlawed, it became even more important to be able to study nuclear reactions in the laboratory (Princeton Plasma Lab), through accelerators that smashed elemental particles into each other (Brookhaven, SLAC, Thomas Jefferson Lab, Fermilab), and using supercomputers that could simulate the reactions (many of the labs).

As Cold War tensions eased and funding went down, the labs (with their thousands of scientists and engineers) looked for other projects that they could work on. Some labs focused on one or another of these scientific aspects of weapons production and some developed new topics to research (medical radiation, cryptography, etc.) and began seeking contracts from other agencies of the US government and from commercial businesses. An additional incentive for labs to diversify was provided by the cap enacted by Congress in 1964 on the funds that the Department of Defense could obligate to National Labs (Dale and Moy, 2000).

Thus by the 1960s the focus of the labs broadened beyond nuclear research to encompass a wide range of scientific fields, including renewable energy, environmental science, and advanced computing. Today, these labs collaborate with universities, industries, and government agencies to drive technological advancements and stimulate economic growth. The labs

[4]See the history pages for each of the labs. The three labs that didn't follow this course are:

- National Energy Technology Laboratory (NETL), which grew out of research on coal mines and petroleum production,
- National Renewable Energy Laboratory (NREL), which was established to focus on wind, solar, etc.
- Lawrence Berkeley Laboratory, which was founded in 1931 to study radiation using a cyclotron (developed by Ernest Lawrence) and then did some of the theoretical work to develop the atomic bomb as well as developing medical radiation.

developed the technology for CDs and DVDs, rechargeable lithium batteries, the cathodes most used in electric vehicle batteries, and many, many other technologies. The research conducted at Argonne and other national labs has had far-reaching implications, affecting not only national security but also contributing significantly to economic growth and public health. Today an important goal of the labs is to serve as hubs of innovation, fostering the development of advanced technologies and creating high-quality jobs in the surrounding regions as well as in the nation as a whole.

These labs vary in size (both by location and over time). Typically, however, they represent significant infusions of resources into their local communities, especially when they were first established. Because of the danger and classified nature of their early work, the government often sought to place them in isolated locations. As their missions broadened, most of the labs have become more public-facing and are important parts of their local innovation ecosystems (though some argue that the labs are excessively "cloistered" (Jaffe and Lerner, 2001; Tartari and Stern, 2021; Dale and Moy, 2000).)

One example is Argonne National Laboratory, located in Lemont, Illinois. Established in 1946 as the first national lab in the U.S., Argonne evolved from the University of Chicago's Metallurgical Laboratory, which was instrumental in the Manhattan Project. Initially, its mission centered on developing nuclear reactors for both peaceful and military purposes. Argonne's rural location was specifically chosen to facilitate safe testing of nuclear technologies (previously the testing was done underneath the University of Chicago's football stadium, which was located in a populated area). Argonne had 2, 727 employees in 1957, 2.5% of the total workforce of DuPage County. In 2022, it had 3,723 full-time employees and a budget of $1.1 billion.[5]

Over the years, Argonne has expanded its research scope to include a broad array of scientific disciplines, such as physics, chemistry, biology, and engineering. Its collaborations with universities, industries, and government agencies have been instrumental in solving complex scientific challenges and driving innovations in various fields.

3 Literature Review

The economic impact of research institutions, particularly government-funded ones, has been examined extensively. Jaffe (1989) pioneered work on the spillover effects of research universities, showing that proximity to such institutions leads to higher innovation outcomes, especially patents. Later studies extended Jaffe's findings by examining not just the impact

[5] "Argonne National Laboratory Annual Report: 1958" click here for link p. 99; "Argonne: By the Numbers" click here for link.

on patent activity but also the broader economic consequences, such as wage growth and firm formation. Feldman and Kelley (2006) found that areas with strong research institutions tend to see increases in high-wage, high-skill job creation, as the proximity to cutting-edge research fosters entrepreneurial ventures and attracts highly skilled workers. Tartari and Stern (2021) found that increased research budgets for universities caused increases in quality -adjusted entrepreneurship in recent decades. In contrast, increased spending by national labs does not have positive effects on this variable.[6]

Andrews (2023) uses archival data to identify counterfactual locations for universities, and finds that universities are associated with large increases in patenting over time. Other papers use this approach to study educational outcomes ((Russell et al., 2024)) and mobility ((Andrews, 2023)). Similarly, Hausman (2022) looks at the spillover effects of university innovation on growth in establishments and employment. These papers build on a large literature relating universities to innovation and entrepreneurship (Aghion et al., 2009; Furman and MacGarvie, 2007; Hausman, 2022; Kantor and Whalley, 2014).

These studies highlight the economic impact of research institutions through the mechanism of knowledge spillovers. That is, research outputs from these institutions are adopted by nearby firms or inspire local entrepreneurial ventures. Some research has examined these mechanisms directly. Zucker et al. (1998) emphasize the importance of "star scientists" in research institutions, noting that areas with a concentration of highly cited researchers tend to see a spike in patenting activity and venture capital investment. Belenzon and Schankerman (2013) find that university knowledge spillovers are geographically constrained, since "citations to patents decline sharply with distance". Jaffe et al. (2000) find that these citation do reflect communication among inventors in an ecosystem, although the signal is somewhat noisy.

Other studies, such as those by Glaeser et al. (1992), have explored the broader regional economic impacts of public and private research organizations. They find that research institutions tend to increase the availability of skilled labor and capital, which, in turn, stimulates wage growth and innovation. Large-scale government investment in research infrastructure also has local impact by increasing demand for goods and services. Thus the establishment of a national lab may act as a catalyst, creating clusters of firms and industries that benefit from the proximity to cutting-edge research, providing increased employment and income opportunities even to non-scientists.

Research on innovation clusters has tended to focus on metrics related to innovation,

[6] A potential explanation for this finding is that Tartari and Stern (2021) examine all 46 Federally Funded Research and Development Centers. Not all of these FFRDCs focus on scientific research; hence our study will focus on just a subset of FFRDCs that are national labs.

and not much on the distribution of the benefits of that innovation. An exception is Gruber and Johnson (2019), who note that most innovation and also most wealth is concentrated in just a few counties, most located on America's coasts. A largely separate literature looks at the impact of place on levels and distribution of income. For example, Moretti (2004) finds that increases in a city's number of college graduates raises wages for workers with all levels of education in that city. Bartik (2020) argues that policies such as improving local infrastructure and training can improve employment outcomes more effectively and equitably than non-geographically targeted policies.

The next section onwards proceeds to empirically investigate our research questions.

4 Data Sources

Below is a description of the data we use in this version of the paper. Under "future research" below, we describe the wealth of information we have uncovered, but not yet incorporated into our analysis.

Patent Data from HISTPAT: To capture innovation outcomes, we primarily use patent data from the HISTPAT database (downloaded from Harvard's Dataverse), a comprehensive source of historical patent records. We focus on inventors and furthermore leverage information about their location of residence. We then aggregate this data to the county level to create a panel of county-year patents starting in the mid-nineteenth century up till 1975, the last year available in the HISTPAT dataset.

OSTI Archives: This will be supplemented with data directly obtained from a few of the labs themselves, which detail patents that the institutions were actively involved in producing. The Office of Scientific and Technical Information (OSTI) preserves a voluminous archive of reports by some of these labs, some of which list the patents affiliated to their specific lab. By combining HISTPAT data with National Lab-specific records, we can more precisely measure the contribution of the National Labs to local and national innovation. This approach allows us to assess not only the quantity but also the quality of patents associated with the National Labs and the industries they influenced most directly.

Census Data: We also use Census data from IPUMS for three time periods (1940, 1960, 1970) to study changes to individual level wages for those National Labs located in counties with sufficient population to avoid suppression of detailed Census data for disclosure reasons. The advantage of this dataset is its richness at the individual level and the large sample size it provides.

Other sources: We also obtain a wealth of information from multiple archives and books that document information such as proposed alternate sites for a few National Labs, detailed

records on the size and scope of the research centers, and occasionally even information about employees, including their home locations (Crease, 1999; Holl, 1994; Nemett and Adams, 2022; Thayer, 1996). Such detailed documentation provides invaluable context to the establishment and operation of the National Labs, allowing us to explore the regional factors influencing site selection and the impact of these institutions on local populations.

5 Empirical Methods

To assess the impact of the National Lab openings on local innovation outcomes, we will employ synthetic control models to measure the effect of each National Lab's establishment on patenting activity. The patenting data comes from the HISTPAT database, which we already have in hand, supplemented by additional National Lab-specific patent data from archival sources. The synthetic control method allows us to construct a counterfactual region, or "synthetic control," for each National Lab's location, composed of a weighted combination of other regions that did not experience a National Lab opening. By comparing patent activity in the real location to this synthetic control, we can estimate the causal effect of the National Lab's establishment on local innovation. This approach is particularly useful in cases where randomized experiments are not feasible and provides a robust way to isolate the effect of the National Labs from broader national or regional trends in patenting.

5.1 Use of Alternate Proposed Locations as a Control Group

In addition to using standard synthetic controls, we leverage archival data from a variety of sources (Crease, 1999; Holl, 1994; Nemett and Adams, 2022; Thayer, 1996) to incorporate alternate proposed National Lab locations as natural control regions. These alternate sites, which were considered but ultimately not selected for the National Lab, offer a unique opportunity to create a more precise counterfactual (see Table-2 for a select few cases). Since these locations were viable candidates for the National Labs but did not receive them, they share many of the same geographic, economic, and demographic characteristics as the actual National Lab sites. This provides a more valid comparison group than traditional synthetic controls, which are constructed from a weighted average of multiple regions.

By comparing patent activity in the chosen National Lab location to that in the alternate proposed sites, we can better isolate the effects of the National Lab itself from broader regional trends. This approach reduces concerns about selection bias, as both the treated and control areas were part of the same decision-making process. Additionally, the use of alternate proposed locations allows us to assess whether the presence of the National Lab

Table 2: Winner and Runner-up Counties identified for 4 National Labs

Winner County	**Runner-up Counties**
Brookhaven National Laboratory (Suffolk, NY)	Rockland, NY Morris, NJ Westchester, NY Fairfield, CT Bergen, NJ Mercer, NJ Middlesex, MA Monmouth, NJ Orange, NY New Haven, CT
Idaho National Laboratory (Bonneville, ID)	Valley, ID
National Renewable Energy Laboratory (Jefferson, CO)	El Paso, CO
Pacific Northwest National Laboratory (Benton, WA)	Douglas, WA Deschutes, OR Wasco, OR Shasta, CA Riverside, CA Kootenai, ID

had a unique impact or if similar outcomes might have occurred in any of the proposed regions.

5.2 Disaggregating Patent Counts

We further refine our analysis by disaggregating patent counts into those directly affiliated with the National Lab and those independently produced by firms and individuals in the surrounding region. Using archival materials from OSTI.gov and National Lab records, we identify patents directly connected to National Lab research or personnel. This will allow us to distinguish between patents that are direct products of National Lab activity and those that reflect broader spillover effects on the local innovation ecosystem. This level of granularity enhances our understanding of the direct versus indirect contributions of National Lab to local patenting.

6 Results

Section 6.1 reports results for patent grants using the synthetic control method, and section 6.2 uses a diff-in-diff to study changes in wage growth.

6.1 Synthetic Control Method

We begin this section with results using the synthetic control method. The unit of observation we consider is a county, and a county is considered treated if it has a National Lab established in it at some point in time. Rather than pool the National Labs into a common dummy variable, we run individual analyses for each treated county. This approach, besides methodological ease (Abadie, 2021), also gives us a more transparent and easily interpretable picture of each treated county.

The data for the synthetic method comes from HistPat (downloaded from Harvard's Dataverse). This dataset includes the names of all inventors and the city/town of their residence on an annual basis. This allows us to aggregate the total number of patents granted within a county for a given year. To create the donor pool, we use all non-treated counties in the same state as the treated one. The idea here is to control for other state-level unobserved characteristics or observed but hard to account for state-level policies. For all results to follow we use pre-treatment outcomes at two-year intervals as the predictors. This has been proposed as a good candidate for choosing the covariates (Kaul et al., 2022). Because the HistPat data ends at 1975 we cannot include two labs that were founded after that year, namely, Thomas Jefferson National Accelerator Lab founded in 1984, and National Renewable Energy Lab founded in 1977. Another two were founded in counties that already had a National Lab established earlier (Fermi National Accelerator Lab in DuPage county that already had Argonne National Lab, and Lawrence Livermore National Lab in Alameda county that already had Lawrence Berkeley National Lab).

As Figures 1 and 2 show, for nine out of the twelve cases we consider, the opening of a National Lab leads to an increase in the number of patents granted to inventors in that county relative to the synthetic control counties. The three National Labs that don't show any effect are Ames National Lab, Savannah River National Lab, and Lawrence Berkeley National Laboratory. The results are perhaps not surprising for Ames National Lab owing to its scale of operation. In fact the lab comprises one building located at Iowa State University with a staff of a few hundred people in its early years. Nonetheless a quick eyeballing does suggests a gentle set of spikes in the post years relative to the synthetic control. For Aiken county, South Carolina, where the Savannah River National Lab was founded in 1952, there clearly is an upward trend in the number of patents, though this is overshadowed by the fact that the synthetic control also trends upwards post lab establishment. The puzzling case is for Alameda county where the Lawrence Berkeley National Laboratory located at the University of California Berkeley, Alameda county, CA; these results are probably explained by the lab's proximity to other innovation hubs (such as the nascent Silicon Valley).

Argonne National Laboratory: We now turn to a special case of a National Lab for

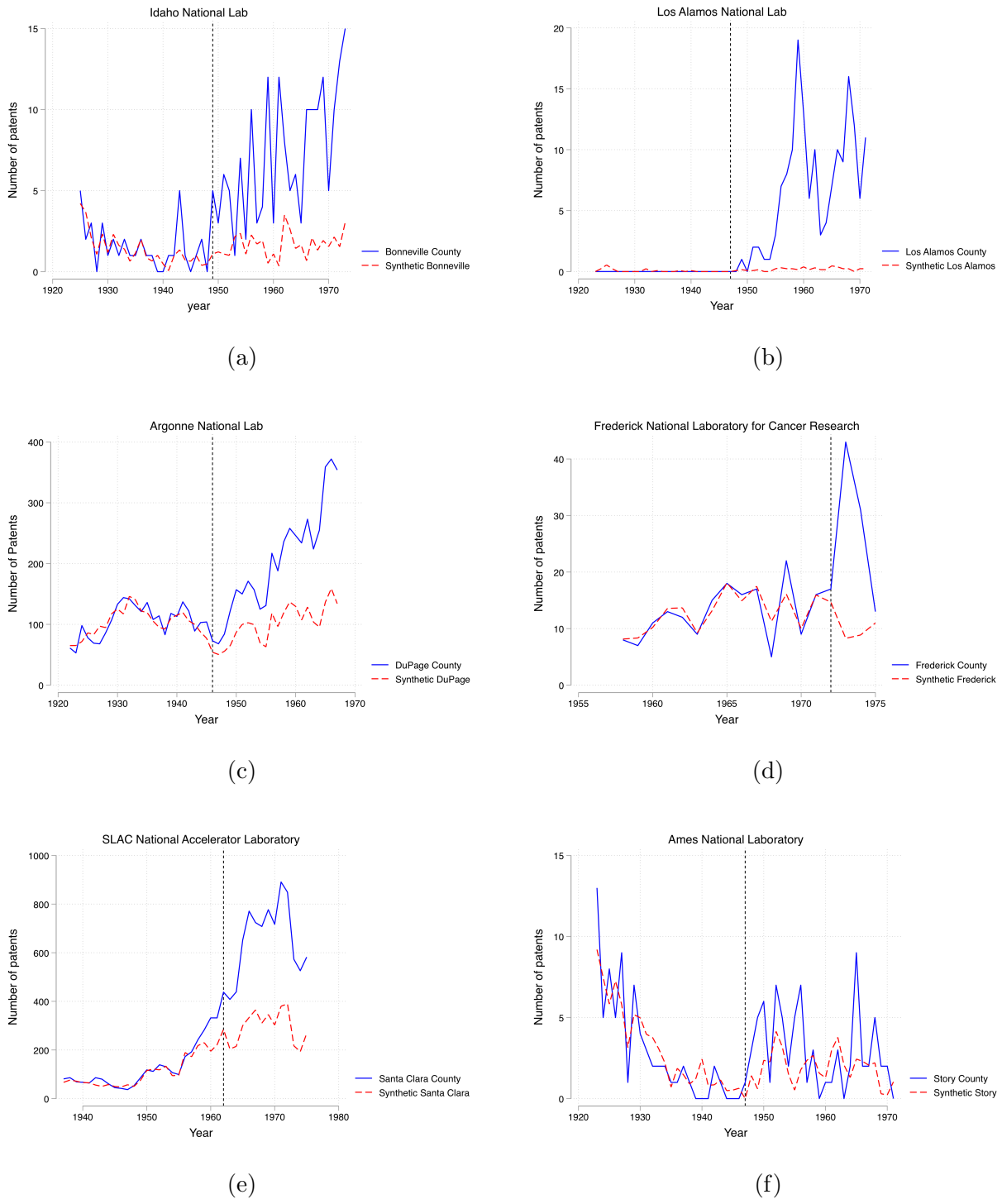


Figure 1: Synthetic Control plots for number of patents. Blue (solid) lines indicate patents for treated counties and red (dashed) for the synthetic control counties. The vertical dashed line in each sub-graph indicates the founding year of the respective National Labs.

which we can decompose the patents. Using archival information from osti.gov, we were

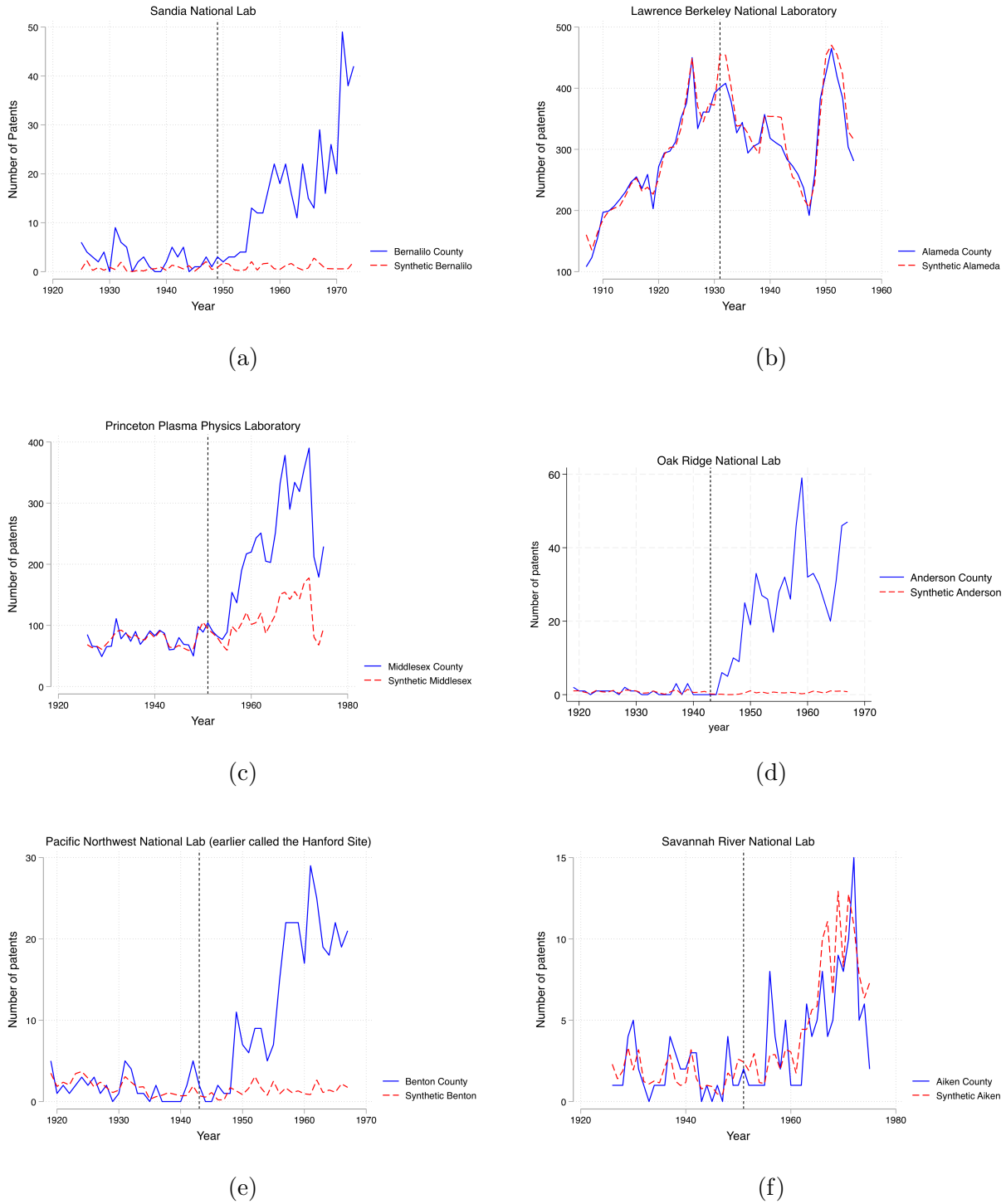


Figure 2: Synthetic Control plots for number of patents. Blue (solid) lines indicate patents for treated counties and red (dashed) for the synthetic control counties. The vertical dashed line in each sub-graph indicates the founding year of the respective National Labs.

able to obtain a list of all patents affiliated to inventors working at Argonne National Lab in DuPage County, IL, from its founding up till 1976. We can then use this information to further isolate non-Argonne affiliated patents. Figure-3 shows these trends.

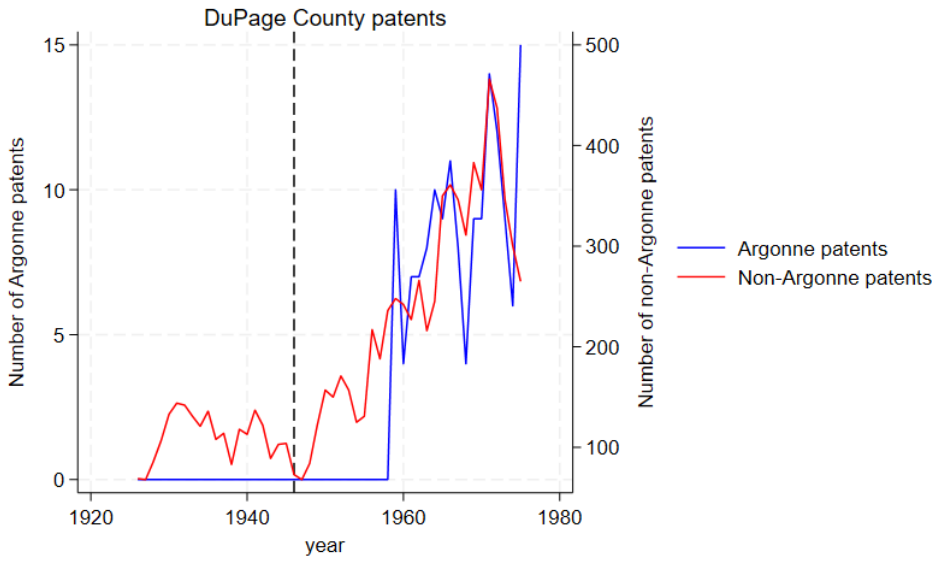


Figure 3: Dis-aggregated patents in DuPage county, Illinois.

The first Argonne patent was granted in 1958, thereafter their number of patents gradually rose (blue line, left scale). This pattern correlates with movements of non-Argonne patents (red line, right scale). Though we cannot completely rule out a third factor driving both time series, the movements are nonetheless suggestive.

6.2 Difference-in-Differences

Next we leverage information on a few identified runner-up counties that were considered as potential sites to the National Labs. Prior to establishment the concerned department (eg., Energy, Defense) sets up a site selection committee tasked with identifying suitable locations. Using archival sources and a number of books we have so far identified runner-up locations for four National Labs. Table-2 gives more details, and Figure-A1 in the appendix gives visual descriptions of these locations.

Using this information we look at a particular National Lab—the Brookhaven National Lab in Suffolk County, NY (founded in 1947)—for which we have county information available for both the pre- and post treatment periods. Due to certain census restrictions the location information is not available for the other labs to run the diff-in-diff specification

outlined below.[7] For Brookhaven, out of the ten runner-ups we include nine of them in the control groups since Middlesex County, NJ established its own National Lab in 1951 (the Princeton Plasma Physics Lab).

Given that the location of these National Lab investments were likely not random, by leveraging the runner-up locations as a control group, we can be more assured of taking care of unobserved characteristics that might drive outcomes over time. Using the nine runner-up counties as the control group we run a simple regression as follows:

$$y_{ict} = \alpha_c + \gamma_t + \sum_t \beta_t \ (\text{Treat}_c \times \text{Year}_t) + X_{it} + \epsilon_{ict} \tag{1}$$

where y_{ict} is the log of real wages for individual i in county c at time t, and X_{it} is a vector of individual characteristics, including age, age squared, education level, and marital status. We include county fixed effects α_c, which capture time-invariant characteristics of each county, and year fixed effects γ_t, which capture common shocks or trends over time that affect all counties equally.

The interaction term, $\text{Treat}_c \times \text{Year}_t$, is the main focus of our analysis. Here, Year_t is an indicator variable that equals 1 for the post-treatment periods (1960 and 1970) and 0 for the pre-treatment period (1940). Treat_c is an indicator variable that equals 1 for Suffolk county, the treated county where the National Lab was established, and 0 for the runner-up counties that did not receive the lab. The coefficients of interest are β_{1960} and β_{1970}, which capture the average treatment effect on the treated (ATT) in each post-treatment year.

The control variables X_{it} include individual-level demographic and socioeconomic characteristics that could influence wages. These controls help account for any differences in wages due to factors such as the individual's age, education, and marital status, ensuring that our estimates of the treatment effect are not biased by such characteristics.

The error term ϵ_{ict} captures all unobserved factors affecting wages that are not explained by the included variables.

The main goal of this specification is to isolate the causal impact of the National Lab's establishment on real wages in Suffolk county, relative to the runner-up counties. By including both county and year fixed effects, we control for unobserved heterogeneity across counties and for common temporal shocks, ensuring that the remaining variation used to estimate the treatment effect comes from within-county wage changes over time.

From Table-3, we see that the coefficients β_{1960} and β_{1970} are both positive and statistically

[7] The Census suppresses location information for smaller counties to preserve the anonymity of individuals enumerated.

Table 3: Real Log Wages (Suffolk County vs. Runner-up Counties)

Variable	**(1)**	**(2)**
Treat_c x 1960	0.363***	0.328***
	(0.056)	(0.059)
Treat_c x 1970	0.412***	0.364***
	(0.054)	(0.056)
Age		0.116***
		(0.012)
Age squared		−0.001***
		(0.000)
Schooling years		0.032***
		(0.003)
Other controls	No	Yes
Adjusted R^2	0.036	0.046
Observations	1,545,051	1,545,051

Notes: Standard errors in parenthesis are clustered at the county level. The dependent variable is log real wage in 1999 dollars and Treat_c equals 1 if county c is Suffolk, and 0 otherwise. The unit of observation in both columns is an individual between 18 and 65 years of age and in the labor force. *** p<0.01, ** p<0.05, * p<0.1

significant, indicating a large increase in real wages for individuals in Suffolk county post-treatment compared to the runner-up counties. Specifically, real wages in Suffolk county increased by approximately 35% more than in the runner-up counties between 1940 and the post-treatment periods of 1960 and 1970. This corresponds to an increase in the average annual growth rate of 1.34% throughout the 30-year period.

Column 1 of Table-3 presents the results of equation (1) without the control variables X_{it}, while Column 2 presents the full specification, including the individual-level controls. The inclusion of the control vector does not significantly change the treatment effect, suggesting that the impact of the National Lab establishment on real wages in Suffolk county is robust to the inclusion of demographic and socioeconomic controls.

7 Implications and Future Research

In general, we find evidence that the presence of a national lab in a county leads to increased patenting in that county compared to counterfactual counties, and that not all of the increased patenting is due to the efforts of lab employees, suggesting the potential of knowledge spillovers from the lab to other inventors in the county.

Our results so far may underestimate these impacts, because lab employees who live in neighboring counties will have their patents attributed instead to counterfactual counties in our synthetic control analysis. Our results may also overestimate the impacts if inventors disproportionately locate in a county for reasons unrelated to the lab's presence. Our next steps are to address each of these issues. We will use patent data that allows us to determine inventors' affiliations (as we did for Argonne), and we will make use of data on patent citations, as a proxy for whether knowledge spillovers did occur, i.e., whether local inventors did communicate with lab personnel (Jaffe et al., 2000).

We have only scratched the surface. In the next few weeks, we plan to do the following:

- Using archival data, we can identify more probable "runner up" locations. We could then compare pre- and post-National Lab trends for these locations.
- Importantly, dozens of National Labs were phased out, closed, or decertified. Again, assigning counterfactual locations, the same exercise could be constructed in reverse.
- Identify impacts of policy changes that might affect the extent of knowledge transfer from labs, such as the cap enacted in 1964 on the level of expenditure by the Department of Defense at National Labs that caused labs to broaden their research efforts and client base (Dale and Moy, 2000), or the 1980s changes in technology transfer policies affecting the national labs (Jaffe and Lerner, 2001).
- Add additional research-focused government labs, such as NASA's research centers.

  We also plan to do much more research on the socio-economic impacts of labs, by using individual-level Census data, as described below.

7.1 Additional Data Sources

Detailed locations of Lab employees: Galt (1968), in his dissertation published in 1968, collected locations of all employees of Argonne National Lab in 1966. At the time the laboratory employed over 4600 people who were scattered in the area around the lab. This rich information provides us with micro-level variation at the intensive margin of Argonne's presence in the area. Figure-4 gives us a picture of the granularity of the data.

Measuring Business Formation Using Digitized Business Directories: In addition to analyzing patenting activity and wage changes, we will examine the impact of National Lab openings on local business formation using digitized business directories, such as City Data Books. These directories provide historical records of businesses, including details on business registrations, corporations, and sometimes even information on the size and

TABLE 4

CROSS TABULATION OF AREA OF RESIDENCE BY OCCUPATION FOR WHITE MALES

Community Area	Occupation								Row Total
	Managers	Profes-sionals	Tech-nicians	Clericals	Craftsmen	Operatives	Laborers	Service Workers	
Chicago Community Areas									
Rogers Park			2						2
West Ridge		2							2
Uptown		2	3					1	6
North Center		1							1
Lake View					1				1
Lincoln Park		2							2
Near North Side	1	11							12
Edison Park			1						1
Norwood Park		3							3
Jefferson Park					1	1			2
North Park		1			1				2
Albany Park			1	1					2
Portage Park		1	3						4
Irving Park	1							1	2
Dunning		2	1			1			4
Belmont Cragin			2		1				3
Logan Square		2							2
Humboldt Park		1							1
West Town		1	2		2				5
Austin		4	1						5
North Lawndale			1						1

Figure 4: Excerpt from Galt (1968) detailing employee residence of Argonne National Lab, Illinois.

type of businesses. By analyzing these records, we can track the evolution of local business ecosystems before and after the establishment of a National Lab. A starting point for us is the County and City Data Book, 1947-1977 (ICPSR) that has annual data on a number of outcomes we feel are relevant for our research questions.

Using pre- and post-National Lab data from City Data Books, we will measure changes in the rate of new business formations, closures, and expansions in the areas surrounding the National Lab. This approach will help us assess whether the presence of a National Lab stimulates local entrepreneurship and business growth, particularly in industries related to research and development, or if business activity remains relatively unaffected.

The longitudinal nature of these directories allows us to capture trends over time, providing a continuous picture of business dynamics. By comparing business formation rates in the years leading up to a National Lab opening with those in the years following, we can identify any significant shifts in economic activity. Furthermore, we can compare these trends

with similar regions that were not selected as National Lab sites (using synthetic control or alternate proposed locations) to ensure that any observed changes are attributable to the National Lab itself rather than broader economic conditions.

We can also disaggregate the data to focus on different types of businesses (e.g., tech startups, manufacturing firms, service providers) and determine which sectors were most affected by the presence of a National Lab. This will give us a clearer understanding of how the economic impact of National Labs is distributed across industries and whether certain sectors disproportionately benefit from their proximity to high-level research institutions.

By integrating business directory data with our other sources, we will develop a comprehensive view of the economic changes in National Lab regions, examining not just innovation and wage effects, but also broader indicators of economic vibrancy, such as new business creation and industrial diversification.

Runner-up counties for the other labs: We acknowledge that, at this stage, only a limited number of runner-up counties have been identified for our analysis. However, we are actively working to expand this list, aiming to include most of the remaining National Labs. The identification of additional runner-up counties is a crucial next step, as it will allow for a more comprehensive comparison and further validate our findings.

By increasing the size of the runner-up pool, we can better account for regional variations and strengthen the credibility of our results. A larger dataset will not only bolster the robustness of our synthetic control estimates but also provide a more nuanced understanding of the long-term impacts of National Labs on local economies. As we continue this work, we are confident that the inclusion of more runner-up counties will significantly enhance the overall contribution of this paper to the literature on innovation clusters and regional development.

8 Conclusion

We report some early results from an NSF-funded project on understanding the origins and outcomes of technology and science-based regional ecosystems. We look at the impacts of large federal investments in National Laboratories on both innovation-related and socioeconomic outcomes in surrounding geographies. Although most of the labs were founded with national security considerations in mind and thus did not initially seek strong connections with local communities, we find evidence consistent with a hypothesis that they have had significant economic impact.

Using both synthetic controls and "runner-up" locations, we find that the establishment of most labs led to significantly more patenting in their host counties, compared to similar

regions and that these differences grew over time. We conducted some case studies that yielded additional suggestive evidence of spillovers. In the one case (Argonne) where we had a list of employee patents, we found that most of the increased patenting was not due to patents received by Argonne employees, suggesting the possibility of spillovers. Our efforts to look at socioeconomic outcomes are more nascent. In the one case in which data for both winner and runner-up counties was not suppressed by Census we found a significant, sustained increase in mean wages in the winning county (Suffolk County, Brookhaven National Lab) compared to runner-up counties.

Our evidence for innovation and socio-economic spillovers so far is at best suggestive. We can work to rule out alternative explanations for our findings by incorporating data on patent citations, patent assignment, and rich individual-level data available in national lab archives, in older Censuses as well as the County and City Data Book 1947-1977 (ICPSR). Finally, we intend to use digitized City Book data to track the impact of the labs on business formation.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of economic literature*, 59(2):391–425.

Aghion, P., Boustan, L., Hoxby, C., and Vandenbussche, J. (2009). The causal impact of education on economic growth: Evidence from u.s. *Brookings Papers on Economic Activity*, pages 1–45.

Andrews, M. J. (2023). How do institutions of higher education affect local invention? evidence from the establishment of us colleges. *American Economic Journal: Economic Policy*, 15(2):1–41.

Bartik, T. J. (2020). Using place-based jobs policies to help distressed communities. *Journal of Economic Perspectives*, 34(3):99–127.

Belenzon, S. and Schankerman, M. (2013). Spreading the word: Geography, policy, and knowledge spillovers. *Review of Economics and Statistics*, 95(3):884–903.

Busso, M., Gregory, J., and Kline, P. (2013). Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, 103(2):897–947.

Crease, R. P. (1999). *Making physics: a biography of Brookhaven National Laboratory, 1946-1972*. University of Chicago Press.

CRS (2020). Federally funded research and development centers (ffrdcs): Background and issues for congress. Technical report, CRS (Congressional Research Service).

Dale, B. C. and Moy, T. D. (2000). The rise of federally funded research and development centers. Technical report, Sandia National Labs., Albuquerque, NM, and Livermore, CA (US).

Feldman, M. P. and Kelley, M. R. (2006). The ex ante assessment of knowledge spillovers: Government r&d policy, economic incentives and private firm behavior. *Research policy*, 35(10):1509–1521.

Furman, J. L. and MacGarvie, M. J. (2007). Academic science and the birth of industrial research laboratories in the u.s. pharmaceutical industry. *The Journal of Economic Behavior and Organization*, 63(4):756–776.

Galt, J. E. (1968). *The Residential Distribution of the Employees of Argonne National Laboratory: Patterns and Implications*. PhD thesis, The University of Chicago.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in cities. *Journal of political economy*, 100(6):1126–1152.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Gruber, J. and Johnson, S. (2019). *Jump-starting America: How breakthrough science can revive economic growth and the American dream*. Hachette UK.

Hausman, N. (2022). University innovation and local economic growth. *Review of Economics and Statistics*, 104(4):718–735.

Holl, J. M. (1994). The national reactor testing station: The atomic energy commission in idaho, 1949-1962. *The Pacific Northwest Quarterly*, 85(1):15–24.

Jaffe, A. B. (1989). Real effects of academic research. *The American economic review*, pages 957–970.

Jaffe, A. B. and Lerner, J. (2001). Reinventing public r&d: Patent policy and the commercialization of national laboratory technologies. *RAND Journal of Economics*, pages 167–198.

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, 90(2):215–218.

Kantor, S. and Whalley, A. (2014). Knowledge spillovers from research universities: Evidence from endowment value shocks. *Review of Economics and Statistics*, 96(1):171–188.

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2022). Standard synthetic control methods: The case of using all preintervention outcomes together with covariates. *Journal of Business & Economic Statistics*, 40(3):1362–1376.

Mazzucato, M. (2021). *Mission Economy: A Moonshot Guide to Changing Capitalism*. Allen Lane.

Moretti, E. (2004). Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review*, 94(3):656–690.

Murphree, M. and Breznitz, D. (2020). *The Run of the Red Queen: Government, Innovation, Globalization, and Economic Growth in China*. Yale University Press.

Nemett, A. and Adams, K. B. (2022). National renewable energy laboratory history: 1977-2016. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

Russell, L. C., Yu, L., and Andrews, M. J. (2024). Higher education and local educational attainment: evidence from the establishment of us colleges. *Review of Economics and Statistics*, 106(4):1146–1156.

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press.

Slattery, C. (2022). Bidding for firms: Subsidy competition in the us. *Available at SSRN 3250356*.

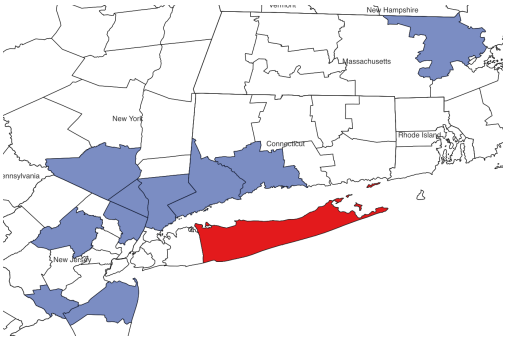
Tartari, V. and Stern, S. (2021). More than an ivory tower: The impact of research institutions on the quantity and quality of entrepreneurship. Technical report, National Bureau of Economic Research.

Thayer, H. (1996). Management of the hanford engineer works in world war ii: How the corps, dupont and the metallurgical laboratory fast tracked the original plutonium works. American Society of Civil Engineers.

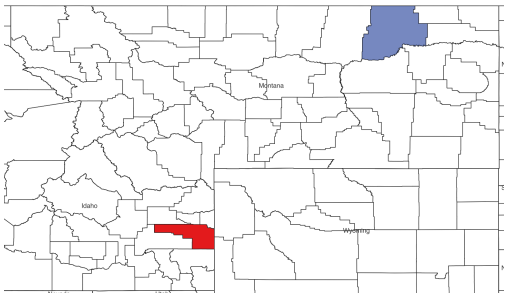
Zucker, L. G., Darby, M. R., and Armstrong, J. (1998). Geographically localized knowledge: spillovers or markets? *Economic inquiry*, 36(1):65–86.

A Appendix

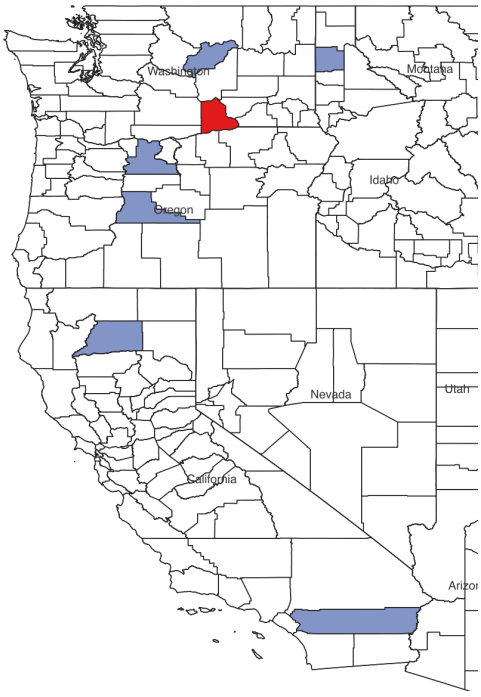
A.1 Map of runner-up counties for a select few National Labs



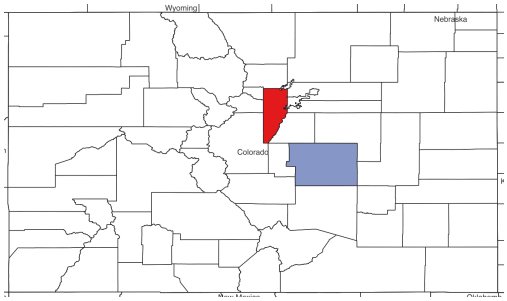
(a) Brookhaven National Lab



(b) Idaho National Lab



(c) Pacific Northwest National Lab



(d) IREL National Lab

Figure A.1: Red counties indicate actual lab location, and blue ones are the runner up counties.