Trends in and Influence of Regional Federally Funded Research and Development in the US

by

Shreyas Gadgin Matha

Submitted to the Institute for Data, Systems, and Society
on May 23, 2018, in partial fulfillment of the requirements for the degree of
Master of Science in Technology and Policy

Abstract

Over the last few decades, although US gross domestic spending on Research and Development (R&D) as a percentage of GDP has risen from around 2.27% in 1981 to 2.74% in 2016, federal funding for R&D has fallen steadily, from 1.19% to 0.81% over the same period. These changes reflect a broader shift in the US from a government-driven R&D model to a business-driven model. Towards the goal of identifying the regional economic impacts of federally funded R&D, I first build on previous work to develop a method to obtain federal funding for R&D at granular geographic levels using Natural Language Processing (NLP) methods to automatically classify open data on federal contracts and grants as R&D or non-R&D awards. This method results in a 95% accuracy rate in classifying federal awards, and covers 56% of US federal R&D obligations made in the year 2016. As underreporting issues in the data source are addressed, this method will yield higher coverage rates, thus creating a unique dataset that affords opportunities to study the regional impacts of federally funded R&D. Next, I adapt Hausman, N. (2012). University Innovation, Local Economic Growth, and Entrepreneurship to identify the employment-generation effects of federally funded university R&D and compare impacts of overall R&D funding to the employment-generation arising from R&D funding provided to specific academic disciplines. I find that the employment-generation effects of federally funded computer science R&D are significant and much more pronounced than the corresponding effects of overall federally funded university R&D.

Thesis Supervisor: Jonathan Gruber
Title: Ford Professor of Economics
Department of Economics
Acknowledgments

I am extremely grateful to my advisor Jonathan Gruber for his guidance and support throughout my time at MIT. Jon’s energy and enthusiasm for this research topic has been very infectious, and provided the motivation I needed for this project. I would also like to thank Simon Johnson for providing me the advice and support I needed to work on this project. Assisting Jon and Simon with their book has been inspirational - their passion for bringing their expertise as academics to effect real change through public service will be a goal I will always aspire to.

I would also like to thank a few other faculty members at MIT and elsewhere, all of whom I look up to as role models, and without whose help I would not be here today. David Autor, who is not only the kindest, but also, by far, the most charismatic academic I’ve met. Michael Greenstone, Rohini Pande, Nicholas Ryan and Anant Sudarshan who provided me the opportunity to work with them on a fantastic research team at J-PAL, and then provided me the support I needed to make it to MIT. All my teachers at MIT have been incredibly thoughtful and invested in my learning, and I’m deeply indebted to them.

My parents have been rocks in my life that I can always lean on for all kinds of support. Without their tireless sacrifices, I would not have been able to come to MIT. As I grow up, I can only wonder in awe at how they raised us while juggling their myriad responsibilities. Darshan is the best little brother I could’ve asked for, even if he’s halfway across the world. His company and support through the years have had an immense impact on my personality, and I hope he flourishes into a kind, generous and successful human being. My girlfriend Nikita has been like family to me for the last two years - she has helped me settle into this new country and provided for me a home away from home.

My closest friends in Cambridge - Sanjay, Shoan, Vinay and Chetana - thank you for being there for me through the highs and lows of MIT and for making this place fun for me. My friends at TPP, thank you for going on this journey together with me, and I sincerely hope we stay in touch to the best of our abilities. I’ve been inspired by your passion for social change, and I have no doubt that you will all push boundaries and help make the world a better place. A big thanks also to friends and family around the world for their love.

I’m grateful to everyone who facilitated for me the opportunity to study at MIT - this is a special place filled with accomplished, kind and incredibly smart people. Lastly, I hope I’m always as fortunate find the strength and opportunities to keep pursuing my goals.
Contents

List of Figures 9
List of Tables 11

1 Introduction 13
  1.1 R&D as a Public Good 14
  1.2 R&D Investments and Economic Growth 14
  1.3 Government R&D and Private R&D 15
  1.4 Structure 16

2 Trends in Government Funding for R&D 17
  2.1 Introduction 17
  2.2 Identifying R&D Awards 21
    2.2.1 Data Sources 21
    2.2.2 Methodology 24
    2.2.3 Results 26

3 University R&D 29
  3.1 Introduction 29
  3.2 Case - Pittsburgh, Pennsylvania 32
  3.3 Geographical Trends 33
  3.4 Regional Impacts 35
    3.4.1 Data Sources 35
    3.4.2 Methodology 36
    3.4.3 Results 37
    3.4.4 Future Work 40

4 Conclusion 41
List of Figures

2-1 R&D by Funding Source - Trends ................................................. 18
2-2 University R&D by Funding Source ............................................. 19
2-3 Federal R&D by Type ................................................................. 20
2-4 Federal R&D Funding by Agency .................................................. 21
2-5 USASpending Data Model - Sources and Validation ......................... 23
2-6 Wordcloud - Frequently Occuring Terms in Grant Titles ................. 26
2-7 Confusion Matrix - Naive Bayes Model ........................................ 27

3-1 University Performed R&D ........................................................... 33
3-2 Federally Funded University R&D ................................................ 33
3-3 University Performed R&D by County ......................................... 34
List of Tables

2.1 Summary of Data Sources .................................................... 24
2.2 Most Informative Words in R&D Classification .......................... 26
2.3 Results - R&D Transactions and Projects .................................. 27

3.1 Replication - Hausman (2012) ............................................. 38
3.2 Panel Regressions - University R&D Expenditures ...................... 39
Chapter 1

Introduction

In his famous report titled *Science, the Endless Frontier*, Vannevar Bush presents the critical need for government support for scientific research through increased funding for R&D (particularly at universities), and through supporting policies that facilitate the increased conduct of R&D in the country (Bush, 1945). He further expands upon the particular need for federal R&D - although commercial incentives cater to the demand for applied research, the essential task of creating knowledge for its own sake (basic research) needs government support to realize the full force of the social externalities it creates. The report significantly influenced post-World War II science-policy in the US, and continues to provide purpose for government-financed R&D.

Indeed, economic theory suggests that technological progress lies at the heart of innovation and sustained economic growth (Solow, 1957; Grossman & Helpman, 1994; Swan, 1956). Endogenous growth theories suggest that investments in R&D lead to knowledge spillovers that lead to sustained economic growth (Romer, 1990; Aghion & Howitt, 1992). Innovation has played a central part in the growth story of the US. Several of these important innovations that spurred economic growth in the US have benefited at least in part from federal funding for Research and Development (R&D). Showcase examples of wildly successful federally funded technologies that affect our lives today include breakthrough innovations such as GPS, Internet and Artificial Intelligence (AI). Each of these technologies today support industries worth billions of dollars, and provide employment to thousands.

An important dimension to the impact of R&D on the economy is location - innovation ecosystems such as Silicon Valley and Boston's Route 128 are surrounded by high-tech industries. The knowledge spillovers created by R&D are localized and dynamic (Woodward, Figueiredo, & Guimarães, 2006; Llamas, 2017). Universities provide a continuous supply of high-skill labor, provide technology transfer, facilitate the creation of startups, and attracts innovative firms to the region, thus playing a major role...
in regional economic development in terms of employment creation and economic growth. In addition, the presence of universities affects aspects such as local land values and cultural amenities (Lendel, 2010).

1.1 R&D as a Public Good

The performance of R&D brings with it broad positive externalities, that are often not internalized by private actors. Previous literature has attempted to identify the magnitude and direction of the economic and innovation externalities of R&D conducted by firms, universities and government. Bloom, Schankerman, and Van Reenen (2013) find that the social returns to R&D conducted by firms are at least twice as much as the private returns, even after considering rivalrous behavior among firms (which they term the "business-stealing" effect of R&D).

An additional aspect to this is described in Llamas (2017), who suggests that innovation is a cumulative process in which ideas are built on previous innovation, thus making it harder for traditional models to fully capture the effects of such dynamic spillovers. Using data on patent citation flows across firms, he goes on to suggest that the past R&D conducted by others that serves as a basis for R&D conducted by a firm, contributes at least as much as the firm's own R&D efforts in terms of increasing productivity of the firm.

Woodward et al. (2006) demonstrate that university R&D expenditures are closely associated with locational decisions by high-tech manufacturing firms, and find that the spillover effects of university R&D expenditures can reach up to 145 miles from the university. Lendel (2010) further suggests that prominent universities have a more pronounced impact on their regional economies. I examine in detail the locational aspects of federal R&D in this thesis.

These locational aspects are important for policymakers due to political concerns arising through the supposed redistributinal nature of such investments if they are focused in lagging areas. The question of whether it is worthwhile to invest in lagging or peripheral regions that might have untapped potential instead of metropolitan areas that might experience economies of scale does not have a clear answer Rodriguez-Pose (2005).

1.2 R&D Investments and Economic Growth

The path to achieving technological progress (and thus economic growth) is less clear - Romer (1990) suggest that investments in R&D are associated with productivity increases and economic growth.
Private R&D investment is often driven by short-term excludable returns, and can miss the potential for large benefits created by the public-sector driven mission-oriented approach towards solving grand societal challenges, that has resulted in major innovations in the past. This short-term approach towards R&D funding has been criticized for excessive “financialization” and increasing inequality, and is often associated with the movement away from basic research (Mazzucato, 2013; Mazzucato & Semieniuk, 2017). Further, studies analyzing the time-patterns of government R&D policies have shown that whereas tax credits to stimulate private R&D favor short term benefits, direct government subsidies for R&D, which show little benefits in the short term, have significant and sustained benefits in the longer term (possibly due to the long-term nature of R&D projects selected by government agencies to receive grants) (See Becker (2015) for overview).

Moretti, Steinwender, and Van Reenen (2016) use defense spending shocks to identify cross-national impacts of R&D spending, and determine that foreign governments’ expenditure on R&D is associated with a small rise in productivity in domestic firms. This points towards the broad knowledge spillovers and potential for economic growth that are created through government R&D investments. In addition, international evidence of the influence of R&D grants on private firms’ innovative activity and growth is provided by Bronzini and Iachini (2014), Jaffe and Le (2015) and others.

(Wilson, 2009) takes on the impact of tax credits on R&D investments, and finds that states that provide tax credits are able to divert R&D from other states into their own. Today, states and cities compete to attract high-tech investments and the job they bring with them, by providing large tax incentives and other perks (eg: cheap land and good infrastructure). This shows that private investments in high-tech sectors (which are usually associated with employment generation) are viewed as an important tool for economic development by local governments.

1.3 Government R&D and Private R&D

Becker (2015) provides a systematic review of the literature on government R&D policies and their effects on private R&D investments, by grouping R&D policies broadly into three types: (1) Tax credits and subsidies (2) University R&D support (3) Support provided to “formal R&D cooperations”. In doing so, she concludes that recent literature is more unanimous in finding a significant positive effect of R&D tax credits in increasing private R&D compared to earlier (pre-2000) literature. Similarly, although earlier studies had conflicting results on whether direct R&D subsidies acted as complements to private R&D or crowded-out private R&D, recent studies favor the claim that government R&D adds to, and does not take away from, private R&D (Becker, 2015).

Block (2011) notes that although government-funded research often results in path-breaking and immensely successful technologies, the government usually does not share directly in the gains from the
growth of businesses based on these technologies. They argue that in the face of limited availability of funds for research and development, intense competition between research groups has emerged due to the non-collaborative nature of the innovation process in which the winner-takes-all. R&D projects funded by Defense Advanced Research Projects Agency (DARPA) and NASA, that are essentially "mission-oriented", have not only led to reduced R&D cost for private entities, but have also created markets to fulfill the technology-demands of the missions undertaken (Mazzucato, 2016).

1.4 Structure

In chapter 2, I describe trends in the distribution of government R&D, and use data on federal awards from the open data platform USASpending.gov to develop a new method of identifying R&D awards. In chapter 3, using data on R&D expenditures from NSF, I describe time and geographic trends in university-performed R&D and federally funded university R&D in the US, and estimate their impacts on regional employment levels.
Chapter 2

Trends in Government Funding for R&D

2.1 Introduction

In this section, I describe time and geographic trends in government-financed R&D. Figure 2-1 shows the trends in and composition of R&D expenditures - up until 1975, federal sources provided a majority of all R&D funding. High rates of contribution of government funding for R&D carried forward after World War 2, which demonstrated the importance of innovation for national security and economic growth. However, since 1964, the contribution of government funding to overall R&D has been on a downward trend, and has fallen to around 22.7% as of 2015.

Figure 2-2 from the American Association for the Advancement of Science (AAAS) further illustrates the importance of government funding specifically for university R&D - in 2016, government sources provided near 60% of all funding for university R&D.

Figure 2-3 further explores the dwindling federal support for R&D - the fall is driven primarily by a reduction in developmental R&D. Part of the steep recent drop may be attributable to a changing definition of R&D - as of FY 2017, federal agencies no longer consider "late-stage development, testing, and evaluation programs" to be a part of R&D. This is reflected in Figure 2-4, which shows a particularly sharp drop in 2017 in the share of federal R&D supported by the Department of Defense, which is the major funder of the now-excluded category of R&D. However, even aside from changing definitions, Figure 2-3 shows a drop in federal funding for R&D from over 1.2% in 1976 to around 0.8% in 2016.
This constitutes a drop of around 33%.

Although businesses focus primarily on developmental R&D, which allows companies to bring products from the lab to market, for the first time since World War 2, a steep rise in industry-support for basic research accompanied with a slowdown of government funding reduced the federal share of basic research in 2015 to less than 50% (Cohen, Nelson, & Walsh, 2002). This jump in industry-driven basic research is encouraging and contrary to Bush’s claim that industry would not be able to keep up with the demand for basic research. However, the slowdown in government funding is an alarming trend. By its very definition, the purpose of basic research is not to provide immediate economic payoffs, but to create a foundation of new knowledge on which applied research and developmental R&D can be conducted. The increasing demand by both government and private entities for demonstrable short-term economic impacts have strained funding for basic research.

The changing nature of private R&D adds to the concern. Mazzucato and Semieniuk (2017) point out that private financing has, in recent years, moved away from activities such as R&D, and that the economy as a whole has shifted focus from such "productive" investments to a more "financialized" setup. They argue that this movement is due to a trend of "short-termism", which emphasizes immediate financial gain over long term investments, and that the type and quality of financing of R&D is an
2.1. INTRODUCTION

University R&D Funding by Source

<table>
<thead>
<tr>
<th>Source</th>
<th>Expenditures in Billions, FY 2017 Dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal</td>
<td>$70</td>
</tr>
<tr>
<td>State and Local</td>
<td>$860</td>
</tr>
<tr>
<td>Universities</td>
<td>$60</td>
</tr>
<tr>
<td>Industry</td>
<td>$40</td>
</tr>
<tr>
<td>State and Local</td>
<td>$30</td>
</tr>
<tr>
<td>Federal</td>
<td>$10</td>
</tr>
<tr>
<td>Other</td>
<td>$50</td>
</tr>
</tbody>
</table>


Figure 2-2: University R&D by Funding Source

important factor in determining outcomes.

Using evidence from a quasi-experiment leveraging grant applications to the Small Business Innovation Research (SBIR) program, Howell (2017) demonstrates the effect of federal R&D on private firms - an early stage grant has strong impacts on the probability of receiving private financing through Venture Capital, and on innovative activity (through patents) and firm revenue.

Data on total government-financed R&D is currently only available at the state-level through the NSF survey Federal Funds for Research and Development. However, given the localized benefits of R&D, it would be useful to study geographically granular patterns in the distribution of R&D. Understanding the distribution of government R&D is necessary in order to examine regional knowledge spillovers and economic impacts, similar to the literature on university R&D impacts.

Further, the distribution of government R&D also has political significance - local governments (cities and states) often compete with each other for high-tech firms to be based in their regions, through tax credits and perks such as easy access to infrastructure. The recent example of the highly publicized bids for the location of Amazon’s HQ2 is an example of the stakes involved - the perks on offer went up to $7 billion. These bids are in pursuit of the jobs that are expected to be created through the expansion

---

2.0-1.5-0.. R&D Type (Total, Basic, Applied, Development)

Data Source: NSF Federal Funds for Research and Development

Figure 2-3: Federal R&D by Type

of high-tech firms into the region, and the growth of supporting industries.

With the importance of studying government R&D and its geographic distribution established, I now present a potential method to obtain R&D spending patterns from federal award-level data. This method is adapted from a report by the Information Technology and Innovation Foundation (ITIF), with changes to enable automated classification of award-level data to identify R&D grants (Wu, Nager, & Chuzhin, 2016). Due to current data limitations on USASpending, this method does not provide comprehensive coverage of R&D projects, but with improving data collection, storage and access systems on the platform, this could serve as a useful proof-of-concept of how open data can be leveraged to better understand how government funds are put to use, and their impacts.
2.2. IDENTIFYING R&D AWARDS

In order to identify the geographic distribution of overall federally funded R&D (as against just university-performed federally funded R&D), I propose a new method - using project-level data from two government sources - USASpending.gov, the US government’s flagship open data tool on government financing of projects, and STAR METRIC’s Federal Reporter, which is an NIH and NSF-led effort to compile data on federal investments on R&D. The two datasets are described in detail here:

2.2.1 Data Sources

The US Treasury-managed website USASpending.gov provides data on actions performed by federal agencies towards federal awards. USASpending obtains its data on contracts from the Federal Procurement Data System Next Generation (FPDS-NG) and on grants and financial assistance awards from individual agencies directly (USASpending.gov, 2018).

USASpending contains data on federal awards, including contracts and grants. This contains details on every action taken on a federal award - including modifications, de-obligations (reduction / adjustments...
in funding for a contract), additional funding, and so on. Further, since 2010, data on sub-awards made by contractors has also been made available. Federal agencies are required to report actions taken on awards within 30 days (except for the Department of Defense, which is allowed 90 days). Federal contracts worth more than $3,000 and grants, loans and other assistance awards worth more than $25,000 are required to be reported within the stipulated time limit. This excludes sensitive projects such as confidential projects by the Department of Defense.

The data quality of USASpending has been questioned several times due to underreporting and inconsistencies in comparison to agencies’ records, and the underlying data collection mechanisms have been modified several times to address these concerns (Gerli, 2017; Government Accountability Office, 2014). However, recent upgrades to the USASpending website have brought easier access, increased transparency and better supporting systems. The Digital Accountability and Transparency Act of 2014 (DATA Act) requires the Treasury to make improvements in underreporting and inconsistencies previously identified, and through a variety of measures, improve the quality and usability of data submitted to and reported by USASpending. The Treasury and OMB were required to ensure that the data on USASpending conformed to "government-wide data standards" by Spring 2018 - accordingly, a new version of USASpending, which was in beta since May 2017, was released in Spring 2018, fully replacing the old version (Gerli, 2017; USASpending.gov, 2018).

Despite these improvements, data quality remains a major limitation. According to the same GAO report, award amounts were found to be inconsistent in 3-7% of cases in comparison with agency held data in 2012 (before improvements in underlying data systems) (Government Accountability Office, 2014). However, the data still provides informative trends on geographic distribution of government R&D.

I consider two types of data on USASpending - contracts and grants. The difference between the two, as per the website, is as follows:

A contract is an agreement between the federal government and a recipient to provide goods and services to the government for a fee. Grants are a form of financial assistance, where a federal agency transfers a thing of value (either money or in kind) to the recipient in order for the recipient to carry out activities or projects to benefit the public. (USASpending.gov, 2018)

Given that the two types of data (contracts and grants) are derived from different sources, they are structured differently, and subject to different kinds of data validation. The data sources and validation procedures are described on the Federal Spending Transparency website, which allows for public collaboration to help USASpending meet the open-data standards set by the DATA ACT. Figure 2-6 illustrates the complex process through which data is prepared for public access through USASpending.

---

2 http://fedspendingtransparency.github.io/
2.2. IDENTIFYING R&D AWARDS

This action-level data is converted to contract-level data by grouping on the unique identifiers PIID, ParentAwardID, and AwardingSubTierAgencyCode, and summing up the obligation changes made towards each award for the relevant year over all corresponding actions, as per USASpending.gov (2018).

The federal budget creates budget authority for each agency to spend money on various programs, then these programs enter into obligations (through contracts or grants) and report these obligations to USASpending for federal assistances (including grants) through an Award Submission Portal (ASP) and FPDS-NG for contracts. This is again distinguished from outlays which corresponds to the actual expenditures incurred by agencies indicated by the receipt of payments from agencies to recipients. Further, federal agency budgets are organized into programs recorded in the Catalog of Federal Domestic Assistance (CFDA), which lists all federal programs under which grants and other federal assistance awards are made.

In order to ensure that we have maximum coverage of R&D data while trying to describe geographic and time trends, as in Wu et al. (2016), I supplement the data from USASpending with STAR METRICS data. STAR METRICS offers much higher coverage rates for a selected set of agencies, namely NIH, NSF and NASA. However, this dataset specifically contains only R&D projects. For all other agencies, I
use data from USASpending.

Other literature that use USASpending data includes Lecy and Thornton (2016), who discuss data quality and usability of data from USASpending and use it to match federal financial assistance awards to non-profit financials. The underlying system of data collection for federal financial assistance awards has since moved from the Federal Assistance Awards Database (FAADS) to the ASP which allows USASpending to collect data directly from agencies and mitigate some of the earlier data quality issues. The ASP was scheduled to be replaced once more by the Federal Assistance Broker Submission (FASB) system in accordance with the DATA Act, by Fall 2017, although there is no update on that so far.

2.2.2 Methodology

Each contract on USASpending has an associated Product or Service Code (PSC) category code defined by the FPDS system, which allows for the identification of R&D contracts, as all R&D contracts are defined to have PSC codes beginning with "A". Also, since data from Federal Reporter are for R&D projects only, project data for NSF, NIH and NASA can be used directly. However, note that since R&D grants do not come with a readily available method to categorize them as R&D or non-R&D, Wu et al. (2016) manually identify grants that are R&D related for the years 2014 and 2015, and sum up the corresponding obligations by place of performance. This is the first attempt to use award-level data to obtain geographic patterns in R&D. I build on this approach to address the issue of efficiency and reproducibility - given that there are 22,145 grant awards in the year 2015 alone, manual classification of grants into R&D and non-R&D buckets is an expensive task, especially if this work is to be extended to cover more years, or create a dynamic tool that updates as new data is added on USASpending. In order to address this problem, I use Natural Language Processing and Machine Learning approaches to automatically identify R&D grants. Table 2.1 summarises the data sources and methodologies used to identify R&D data.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Agencies</th>
<th>R&amp;D Categorization Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAR METRICS - Federal Reporter</td>
<td>NSF, NIH, CDC, FDA</td>
<td>Only R&amp;D projects included</td>
</tr>
<tr>
<td>USASpending - Contracts</td>
<td>DOD, DOE, USDA, NASA</td>
<td>Using FPDS &quot;Product or Service Codes&quot;</td>
</tr>
<tr>
<td>USASpending - Grants</td>
<td>DOD, DOE, USDA, NASA</td>
<td>Natural Language Processing and classification</td>
</tr>
</tbody>
</table>

From here on out, I will focus the discussion on identify R&D grants from USASpending, as R&D contracts and Federal Reporter data come pre-classified.
2.2. IDENTIFYING R&D AWARDS

In the grant data, the key variables that allow us to classify grants into R&D or non-R&D buckets are the grant title and description, and the CFDA program number under which the corresponding funding agency awarded the grant. Broadly, the steps involved in identifying R&D grants include:

1. Using CFDA program descriptions to identify programs that are meant solely for R&D or non-R&D activities, and programs that may be ambiguous
2. Creating a labelled dataset of grants labelled as R&D, non-R&D or ambiguous based on the CFDA program under which the grant was made
3. Extraction of relevant features from the grant titles and descriptions using Natural Language Processing
4. Using supervised classification models to learn models to identify R&D and non-R&D grants based on underlying differences in their grant descriptions

The labelled data is split in a 70:30 ratio into training and test sets respectively, to first fit and then evaluate the model. The grant titles and descriptions are then pre-processed to extract features from the text that allow us to better train the machine learning models that will eventually predict whether any given grant is R&D or non-R&D. Pre-processing steps include:

- Special character removal, case conversion
- Word stemming
- Stopword removal

First, special characters and numbers are removed as they do not carry as much information the nature of the grant as the words in the grant. Next, the words are stemmed so as to remove slight differences in the text of words (derivative forms, such as cooking vs cook) that are otherwise similar, by using the popular Porter Stemmer. Lastly, stopwords (or words commonly occurring in the English language, and thus do not carry much information) are removed - eg: words such as a, an, the and so on. Figure ?? illustrates using a wordcloud a few frequently occuring stemmed terms in the corpus.

The next step is to extract relevant features from this text to conduct supervised classification. For this, I use a bag-of-words model, which counts the number of occurrences of each token (which, in this case corresponds to one word) in each grant. Next, to account for the importance of a certain word to the grant that it appears in, a TF-IDF weighting scheme is used. TF-IDF, or Term Frequency - Inverse Document Frequency, provides increased weights to words that appear multiple times within each grant and reduces the weights of terms that appear frequently across the whole database of grants.

Finally, a Multinomial Naive Bayes Classifier is used to fit the model to the text. The 25 words that are most informative in predicting whether a grant is an R&D grant are listed in table 2.2. Once the model is fitted, it is then used to predict the R&D or non-R&D category of grants made under "ambiguous" CFDA programs.
CHAPTER 2. TRENDS IN GOVERNMENT FUNDING FOR R&D

2.2.3 Results

The Naive Bayes Model results in a 95% accuracy rate. Although more sophisticated models could yield higher accuracy rates, the classifications provided by this model provide results that are illustrative. Figure 2-7 shows a confusion matrix of classification performance.

Table 2.3 summarizes the resulting compilation of transactions from USASpending and projects from STAR METRICS, and the coverage rates achieved as a result. The coverage rates are obtained by comparing the total obligations identified from award-level data to R&D obligations by agency obtained from NSF Federal Funding for Research and Development.

The overall coverage rate of 56% is not sufficient to draw inferences on regional impacts of R&D. However, if the improvements in federal awards systems result in decreased underreporting of R&D
2.2. IDENTIFYING R&D AWARDS

![Confusion Matrix - Naive Bayes Model](image)

**Figure 2-7**: Confusion Matrix - Naive Bayes Model

**Table 2.3**: Results - R&D Transactions and Projects

<table>
<thead>
<tr>
<th>Agency</th>
<th>Number of Transactions</th>
<th>% of Obligations Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOD</td>
<td>63295</td>
<td>40.3%</td>
</tr>
<tr>
<td>DOE</td>
<td>8446</td>
<td>36.5%</td>
</tr>
<tr>
<td>NASA</td>
<td>16898</td>
<td>84.9%</td>
</tr>
<tr>
<td>NIH</td>
<td>67174</td>
<td>95.1%</td>
</tr>
<tr>
<td>NSF</td>
<td>12511</td>
<td>76.8%</td>
</tr>
<tr>
<td>USDA</td>
<td>3831</td>
<td>31.8%</td>
</tr>
<tr>
<td>Total</td>
<td>172155</td>
<td>55.98%</td>
</tr>
</tbody>
</table>

awards, coverage rates might improve. Further, this method of classification can be extended to identify grants of any "type" (not just R&D), as long as a sufficiently large labeled dataset can be created to train the model.
Chapter 3

University R&D

3.1 Introduction

Apart from providing an important source of R&D expenditures, universities also act as hubs of regional economic development (Hausman, 2012; Woodward et al., 2006; Kirchhoff, Newbert, Hasan, & Armstrong, 2007; Lendel, 2010). Apart from the research that takes place within universities, universities support and nurture industry, incubate and spawn startups, promote technology transfer to industry, and provide a continuous source of high-skill labor. Universities also carry out a large portion of the country's basic research - this undirected research provides the foundation upon which innovations are then created in the form of applied research (again often at universities). Since the Bayh-Dole Act of 1980, universities have also taken an active role in technology transfer and commercialization, with many providing nurturing environments for entrepreneurship both within the university and for small businesses in their regions. Further, several universities have played pivotal roles in regional economic development, providing local governments with technical expertise to implement policies to attract high-tech firms, and the workforce to sustain them.

The location of universities is of particular significance - their presence creates new firms through innovation and entrepreneurial support, but also attracts existing high-tech firms. The expansion of high-tech firms into a region generates employment, not just within the relevant high-tech industries, but also in supporting industries - Cohen et al. (2002) suggest that university and government-performed R&D positively affects industrial R&D not just high-tech sectors, but also across much of the manufacturing sector. Further, they identify that research papers, reports, conferences, workshop and other activities conducted by universities act as pathways through which such effects on industrial R&D are
realized. Although the economic influence of universities cuts across industries, their region of influence is usually geographically bounded (Woodward et al., 2006; Anselin, Varga, & Acs, 1997).

Anselin et al. (1997) also explicitly investigate the impact of university R&D on local "innovative activity", and find local spillovers through business R&D. (Kirchhoff et al., 2007) highlight that university R&D expenditures are also associated with the formation of new firms in that region and with economic growth. Mature technologies such as Artificial Intelligence or the internet (both of which were developed through university-industry collaboration) have impacts that cut across geographical boundaries and industries. However, in the short term, the spillover effects of university research on innovative activity and economic development seem to be localized.

The presence of universities contributes to the formation of innovation ecosystems which allow for geographic agglomeration in the knowledge space, which is considered an important factor in raising innovative activity. By providing a nurturing environment for R&D ecosystems, universities also cause agglomeration in industrial activity around them, which in itself comes with strong economic benefits (Greenstone, Hornbeck, & Moretti, 2010). Industry-university linkages led to the formation of the Silicon Valley and Boston's Route 128 high-tech sectors. University research parks are another example of innovation ecosystems forming around universities, attracting firms and thus creating employment.

As Llamas (2017) suggests, knowledge spillovers have a dynamic effect, as new ideas are based off of a foundation of previous research - universities thus have a high societal value by conducting openly accessible research which has a large potential for such dynamic knowledge spillovers.

University R&D data is available at the institution-discipline-funding source level. Using this data aggregated to the county level, I address the question of whether government-funded university R&D in particular, is associated with employment. I replicate and adapt work by Hausman (2012), who uses an external shock (passage of Bayh-Dole Act) and variation in the amount of university R&D received before the act was passed to estimate the impact of federally funded university R&D on local employment, using microdata from the Census Bureau's Longitudinal Business Database (LBD). I use data from the Census Bureau's County Business Patterns (CBP) to estimate the association of university expenditure on R&D in one specific field (Computer Science) with industrial employment. I use Computer Science as the selected discipline because of the availability of data through population surveys back to 1976. Specifically, I identify the effect of university R&D in specific disciplines on employment growth in the areas around them. With the explosive growth of the computer science industry in the last few decades, we might expect to find the employment generation effects of computer science university R&D to be more pronounced than the overall employment generation effects of university R&D.

Hausman (2012) does not examine the impact of funding for university R&D in specific disciplines on immediately relevant industries, although she does examine the impact of university patents on
"relevant industries" using a probabilistic concordance developed by experts. I explore whether there exists discipline-specific heterogeneity in the influence of university R&D on employment. Although it would be interesting to examine the effects of discipline-specific university R&D on immediately relevant industries, this is a challenging task due to the large number of masked employment values in the County Business Patterns data. Employment and payroll values are masked with data suppression flags for a majority of county-industries for each year in the CBP data due to confidentiality concerns. The flags indicate employment ranges, instead of specific employment values (e.g., 0-19 employees, 50,000-99,000 employees, etc.)

There have been several attempts at trying to use "clues" in the CBP data based on employment values in establishment size classes, number of establishments, employment in broader industry code classifications, and other parameters to try and narrow the estimates for employment in each county-industry (Isserman & Westervelt, 2006; Register, 2012; Porter, 2003). However, the data products provided by the authors as outputs from these methods are too expensive (Register (2012) reports that these prices per year can vary between $3000 to over $36,000). Although Register (2012) describes more accessible methods to obtain narrower estimates of employment ranges, the employment ranges obtained are still crude. I leave the examination of the validity of these refined ranges in our analyses for future work. Hence, I consider only data at the county-level (across all industries), which is much more complete - over 99% of the counties for the relevant years have exact employment data available.

With county-industry data, future work could include a probabilistic concordance between university academic disciplines and industrial sectors to contrast the impact of R&D investments on both related and unrelated industries. A major limitation of my current study is that there are endogeneity concerns with identifying the association of university discipline-specific R&D and employment growth - for instance, increased university R&D may be invested in places that have a large workforce, bringing reverse causality issues. However, using the method first demonstrated by Hausman (2012), I use an interaction term with the passage of Bayh-Dole to examine the influence of university R&D on regional employment, given the broad and well-established technology transfer and commercialization incentives provided by the Bayh-Dole Act.

The Bayh-Dole and Stevenson-Wynder acts of 1980, along with subsequent amendments, provided strong incentives for commercialization of research. These acts allowed businesses and non-profits, including universities, to retain title to their federally funded inventions, providing incentives for commercialization and distribution of federally funded R&D to the public (Link, Siegel, & Fleet, 2011; Grimaldi, Kenney, Siegel, & Wright, 2011; Shane, 2004).

First, I present a brief case study to illustrate the influence of universities on regional economies. Youtie and Shapira (2008) considers the case of Georgia Institute of Technology (Georgia Tech), to demonstrate the evolving functions of universities - from simply centers of knowledge creation and dissemination to hubs of economic development and regional economic revitalization. Georgia Tech's leadership
complemented the efforts of the state of Georgia in moving its focus from agriculture to manufacturing and finally to high-tech innovation. I then consider the case of Pittsburgh, Pennsylvania to demonstrate the role of its universities in turning around its economies after the fall in the steel industry.

3.2 Case - Pittsburgh, Pennsylvania

This section considers the case of Pittsburgh, Pennsylvania to demonstrate the influence of universities on regional economic development. Pittsburgh's economy, which was heavily based on steel in the 1980s, faced a major setback in the 1980's as the steel industry collapsed, leaving a massive unemployment problem behind. Carnegie Mellon University (CMU) and the University of Pittsburgh (Pitt) spearheaded the shift to a tech-centric economy. Both universities poured millions of dollars into research in technology (Toland, 2013). This led to a spike in businesses surrounding computer software, robotics and biotechnology, primarily working in the sectors of education and health care. Although Pittsburgh's population declined from 1990-2010 from around 370,000 to around 305,000, it has since stabilized, at around 300,000 residents over the period 2010-2016 (Streitfeld, 2009). The number of jobs in Pittsburgh have gone up by around 122,000 (12%) since 1990, driven primarily by the education and health sectors (Econsult Solutions Inc., 2017). The Pittsburgh Council on Higher Education, a consortium of colleges and universities in the Allegheny County area of Pennsylvania, estimates that its colleges and universities generated an economic impact of $8.99 billion and supported more than 70,000 jobs in Pittsburgh during fiscal year 2012-13 (Pittsburgh Council on Higher Education, 2014).

After the slump, local foundations invested heavily in research. CMU and Pitt concentrated their research on software, robotics and health. Today, the education and health sectors are the two fastest growing sectors in Pittsburgh in terms of employment.

Role of CMU: CMU has played a central role in the regional transformation of the Pittsburgh region, by attracting skilled talent, research dollars, entrepreneurs and private investment. Pittsburgh's turnaround was sparked at CMU, particularly through software and robotics. CMU has brought to Pittsburgh a an emerging technology sector - companies such as Google Facebook and Uber have opened offices near the CMU campus. A CMU sponsored-report points out that although the scale of employment opportunities created by high-tech firms do not match up to the manufacturing industries, it does create potential for exports and spin-offs (Econsult Solutions Inc., 2017). The CMU Robotics Institute is also America's largest robotics research center.

Role of Pitt: The University of Pittsburgh Medical Center (UPMC) is the largest healthcare and insurance provider in Pennsylvania. UPMC has invested close to US$2 billion in technological innovation. "UPMC was the birthplace of PACS, one of the earliest medical data archiving systems. The hospital receives petabytes worth of data, an amount that is doubling every 18 months. Keeping up with the data
volume for real-time decision-making has become a priority.” UPMC today hires data scientists and “technologists” to stay at the frontier of medical research.

3.3 Geographical Trends

Figure 3-1 shows that despite its importance for job-creation and economic development, university-performed R&D is inequitably distributed across the US. Further, the geographic distribution of federally funded R&D performed by universities and other higher-education institutions also closely mirrors this trend, as shown in Figure 3-2. In 2015, higher education institutions in the states of California, New York, Maryland, Pennsylvania, Massachusetts and Texas conducted almost half of all university R&D in the US. In addition, universities in these states also incurred almost half of all federally funded university R&D expenditures in the US.

The geographic distribution of university R&D at the county-level (illustrated by Figure 3-3) shows that even within these states, university R&D is concentrated within a few counties that surround major higher education institutions. The top 9 counties constitute over a quarter of all university R&D expenditures in the US.
University–Performed R&D by County – 2015
Source – NSF

Figure 3-3: University Performed R&D by County
3.4 Regional Impacts

3.4.1 Data Sources

I compile data on county-level employment and number of establishments for 1973 to 2016 from the Census Bureau's County Business Patterns. Data for 1973-1985 were collected from ICPSR, and for 1986 onwards from the Census Bureau website. As in Hausman (2012), I limit the years selected for the data to the census years 1977 to 1997. A limitation in the publicly available County Business Patterns data is that the data on employment is censored in less than 1% of cases with masking flags in order to preserve the confidentiality of establishments' data in cases where there are only a small number of establishments in a county. In these cases, a "data suppression flag" indicating the range of employees in the county-industry is available - I approximate the number of employees by taking the mean of the range boundaries.

Data on R&D funding for Computer Science by academic institution comes from the NSF Higher Education R&D Survey, known before 2010 as the NSF Survey of Research and Development Expenditures at Universities and Colleges. The survey is conducted by the National Center for Science and Engineering Statistics (NCSES). The data prior to 1998 comes from a combination of sample surveys for some years and population surveys for others, with data for sample surveys inflated to represent national-level R&D expenditures. The zip codes for each academic institution is available as well - a key variable in identifying regional impacts. I use a zip code - county crosswalk from USPS to aggregate data to the county level to get total federally funded university R&D for computer science for each county. Next, I assume that every county with a centroid within 75 miles of the centroid of the county in which the university is present, is equally exposed to that university's R&D expenditures, as in Hausman (2012). As a result, for each county, I add R&D expenditures from all contributing universities to get a measure of exposure to federal funding for university R&D in computer science. Ultimately, my data has one record for each year-county pair representing the county's exposure to federally funded university computer science R&D, and the sum of its employment levels across all industries.

Data on overall university R&D used to replicate work by Hausman (2012) comes from the NSF Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions, which is a population survey of all federal agencies that grant money to academic institutions for Science and Engineering R&D. Similar to the NSF Higher Education R&D Survey, zip codes for each academic institution is available, and I aggregate the R&D expenditures from all nearby universities to obtain a measure of exposure to university R&D for each county.
3.4.2 Methodology

As Hausman (2012) explains, an ideal identification strategy would involve randomly assigning university R&D to different locations and observing resulting employment changes. In the absence of such a clear identification strategy, however, I run a panel regression with year-fixed effects to understand whether university computer science R&D is associated with total employment. A naive regression model would thus look as follows:

\[ emp_{ct} = \beta_0 + \beta_1 \text{funds}_c + \phi_t + \epsilon_{ct} \] (3.1)

Here, \( emp_{ct} \) is the total employment across all industries in county \( c \) at time \( t \), \( \text{funds}_c \) corresponds to discipline-specific university R&D funding that county \( c \) is exposed to in the years leading up to Bayh-Dole, and \( \phi_t \) correspond to time fixed effects.

The model described in equation 3.1 does not consider endogeneity concerns arising due to the reverse effect of total employment on university R&D. To address this, I consider an identification strategy from Hausman (2012) to use the Bayh-Dole Act of 1980 as an external shock to identify the influence of university computer science R&D. This method assumes that after Bayh-Dole, due to the increased incentives provided for commercialization, counties that were exposed to higher university R&D before Bayh-Dole in a certain high-tech field would have an increased influence on total employment.

\[ emp_{ct} = \beta_0 + \beta_1 \text{funds}_c + \beta_2 \mathbb{1}_{aBD} + \beta_3 \mathbb{1}_{aBD} \times \text{funds}_c + \phi_t + \epsilon_{ct} \] (3.2)

In equation 3.2, \( \mathbb{1}_{aBD} \) corresponds to an indicator which takes a value of 1 if the corresponding year is after 1980 (when the Bayh Dole act was passed), or 0 before 1980. The coefficient \( \beta_3 \) of the interaction term should give us the influence of university R&D on employment.

I first replicate work by Hausman (2012) for overall Science and Engineering R&D to ensure consistency in the overall results and to identify a benchmark for the employment generation estimates. I use NSF data on Science and Engineering R&D pooled for the full set of years 1976-1980 to identify variation in university R&D. I consider overall county employment levels instead of county-industry employment due to the data availability constraints in the County Business Patterns data described earlier.

For some of the early years (before 1998), data for some years come from sample surveys and for others from population surveys. In addition, federal obligations may see ragged spikes due to the nature of federal contract obligations. To account for these issues, in line with Hausman (2012), I pool data for the years 1976-1980 to detect variation across counties in discipline specific university R&D funding, excluding the year 1978 as that year’s survey was a sample survey (NSF, 2016). Further, since
3.4. REGIONAL IMPACTS

Bayh-Dole affects only counties with some exposure to universities, I drop counties that do not have any exposure to discipline-specific university R&D. To ensure comparability with Hausman (2012), I restrict the county business patterns data to the census years 1977, 1982, and so on until 1997 in 5 year intervals for my analysis. Thus, the indicator function $1_{aBD}$ takes 0 for the year 1977, and 1 for all other years.

3.4.3 Results

I first replicate the results from Hausman (2012) as closely as possible given my data constraints. The results from 3.2 with $funds_c$ representing county exposure to overall Science and Engineering R&D, are presented in Table 3.1. These results suggest that a $1 million increase in overall federal funding for university R&D led to the generation of an additional 40.6 workers per county.

Table 3.2 provides the main estimates from equation 3.2. Results in the table suggest that a $1 million increase in federal funding for computer science R&D led to the generation of an additional 710.5 workers per county, and 47 establishments per county. The large difference in the magnitudes between the results in 3.1 and 3.2 indicate that there exists wide heterogeneity in the effect of university R&D on employment generation, based on the discipline being funded.

The influence of university R&D on employment demonstrated by these results are in line with our hypotheses and previous literature. Further, taken in the context of previous literature, these results suggest that the pathways through which university R&D increases employment are important, and that these may vary across academic disciplines.
**Table 3.1: Replication - Hausman (2012)**

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>FedFunding</td>
<td>104.032***</td>
</tr>
<tr>
<td></td>
<td>(16.762)</td>
</tr>
<tr>
<td>aBD</td>
<td>6,479.426***</td>
</tr>
<tr>
<td>FedFunding:aBD</td>
<td>40.639***</td>
</tr>
<tr>
<td></td>
<td>(5.309)</td>
</tr>
<tr>
<td>Constant</td>
<td>8,793.241***</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Observations</td>
<td>120,345</td>
</tr>
<tr>
<td>R²</td>
<td>0.091</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.091</td>
</tr>
<tr>
<td>F Statistic</td>
<td>4,011.400***</td>
</tr>
</tbody>
</table>

**Note:**
1. Robust standard errors clustered at the county-level.
2. FedFunding are in million $
### 3.4. REGIONAL IMPACTS

#### Table 3.2: Panel Regressions - University R&D Expenditures

<table>
<thead>
<tr>
<th></th>
<th>Employment</th>
<th>Log Employment</th>
<th>Establishments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS Year FE</td>
<td>Year FE</td>
<td>Year FE</td>
</tr>
<tr>
<td>DisciplineFedFunding</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>1,496.571***</td>
<td>1,496.268***</td>
<td>0.042***</td>
</tr>
<tr>
<td></td>
<td>(292.084)</td>
<td>(292.101)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>aBD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11,206.930***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(869.602)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DisciplineFedFunding:aBD</td>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>710.750***</td>
<td>710.509***</td>
<td>0.003***</td>
</tr>
<tr>
<td></td>
<td>(143.385)</td>
<td>(143.189)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21,246.940***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,704.082)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year FE</th>
<th>X</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>109,542</td>
<td>109,542</td>
<td>109,542</td>
</tr>
<tr>
<td>R²</td>
<td>0.022</td>
<td>0.021</td>
<td>0.062</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.022</td>
<td>0.021</td>
<td>0.062</td>
</tr>
<tr>
<td>F Statistic</td>
<td>838.422***</td>
<td>1,184.161***</td>
<td>3,624.896***</td>
</tr>
</tbody>
</table>

**Note:**
- *p<0.1; **p<0.05; ***p<0.01
- 1. Robust standard errors clustered at the county-level.
- 2. DisciplineFedFunding are in million $
3.4.4 Future Work

Abramovsky, Harrison, and Simpson (2007) match the location of R&D labs in Great Britain in selected industrial sectors with the presence of related academic departments, and find evidence that some industries showed pronounced co-location of R&D activity and relevant university departments compared to others. Combined with the evidence from this thesis and previous literature that suggests that university R&D generates employment, this lends credence to the idea that discipline-specific R&D funding could have heterogeneous impacts on employment in different industries. Narrowing employment estimates in County Business Patterns data through methods described in Isserman and Westervelt (2006) and others could allow for the use of industry-fixed effects in estimating the employment-generation effects of discipline-specific university R&D, thus giving more accurate estimates and better.
Chapter 4

Conclusion

In the context of dwindling federal support for R&D as a percent of GDP, it is important to recognize that federal R&D investments play a significant role in the creation of economic opportunities for the US, and can be leveraged as a policy tool to accelerate job-creation. This thesis addresses the regional dimension of the trends in federal R&D and its influence on regional employment generation.

Open data on federal contracts and grants provide opportunities to examine the geographic distribution of federal dollars at a granular level. I first discuss the data sources USASpending.gov and STAR METRIC's Federal Reporter, their evolution and limitations. Although USASpending has often been criticized for its underreporting issues and inconsistencies across agencies in the data reported, the platform has seen major improvements in the last decade due to legislation towards transparency in federal spending. As the requirements and recommendations of the Digital Accountability and Transparency Act (DATA Act) of 2014 are fulfilled, underreporting issues are expected to reduce further, and the platform can provide a rich source of US government spending data.

Identifying the distribution of R&D obligations requires a method to identify R&D contracts and grants from the larger pool of awards. While federal R&D contracts can be identified using their Product or Service Codes (PSC), federal grants do not have any such categorization. I extend Wu et al. (2016) to develop a method to use text from the titles and descriptions of the grants to classify grants as R&D or non-R&D. Using CFDA program descriptions, I first create a labeled dataset to train and test the model. I then pre-process the relevant text and extract informative features to fit a supervised learning model to classify grants.

The method results in high accuracy rates in classification (95%), and allows us to identify R&D awards that cover 56% of federal R&D obligations for the year 2016. There is a wide variation across agencies in
the underreporting issues in the compiled R&D data - while only around 36.5% of DOE R&D obligations are covered, over 95% of NIH R&D obligations are properly identified through this method. Although it might not be possible to identify sub-national trends in R&D obligations due to the non-random nature of the missing R&D awards, the coverage rates are expected to improve as underreporting issues in the data sources are addressed. The stated goal of STAR METRICS is to "create a repository of data and tools that will be useful to assess the impact of federal R&D investments". This effort brings us a step closer to that goal by complementing the STAR METRICS database with additional data sources on federal R&D investments.

Universities attract high-skill labor, facilitate technology transfer and entrepreneurship, and attract high-tech firms to the region. In order to assess the influence on university R&D on regional employment levels, I first replicate work by Hausman (2012) using a different data source (County Business Patterns) to ensure consistency and establish a benchmark for the magnitude of the employment-generation effect. To do this, I use the Bayh-Dole act as an external shock to identify the influence of federal R&D funding on employment. I find that $1 million of additional exposure to federal university R&D funding was associated with 40.6 additional workers per county. However, this does not consider heterogeneity in the influence of funding for various academic disciplines. To address this issue, I repeat the regressions above using variation in computer science R&D, and find that the employment generation effects are much more pronounced - $1 million of addition exposure to federal university R&D in computer science was associated with 710 additional employees and 47 additional establishments. This method suffers from endogeneity concerns due to the fact that R&D funding for other disciplines also affect employment and are also affect by Bayh-Dole. However, it does provide results indicative of wide differences in the nature of the effects of funding for different academic disciplines on employment generation.

This thesis thus demonstrates the importance of studying federal R&D investments at a granular geographic level. The dynamic and localized knowledge spillovers created by federal R&D investments can be leveraged as tools for employment-generation. Although better data sources are required for a systematic study of the influence of overall government R&D on regional economic characteristics, the methods described in this thesis help create a proof-of-concept for how such data might be leveraged for policy relevant to the issue of job creation. In addition to producing high-skilled labor through teaching, this thesis demonstrates that the presence of universities also creates jobs, and that more work is needed in trying to identify the channels through which funding for R&D in various academic disciplines affect employment.
Bibliography


