Research Memo
May 10, 2022

Exploring The American Recovery and Reinvestment Act's Impact on Federal Grant Funding & Renewable Energy Patents

Anthony Chen , Yuehang Chen, Trevor Mattos, Bowen Li, Francisco Pineda, Hanyu Tian, Dingrui Zhang

Abstract

Climate change poses a serious threat to our planet, with growing evidence of severe environmental impacts around the world (Pörtner *et al.* 2022). In an attempt to reduce the emissions linked to climate change, the U.S. government has invested billions of dollars in renewable energy resources and more is arriving with the Infrastructure Investment and Jobs Act of 2021. The literature has demonstrated a link between the Department of Energy-funded intergovernmental grant programs under the *American Recovery and Reinvestment Act* of 2009 (ARRA) and emissions reductions (Lim *et al.* 2021). Yet, the extent - and nature - of basic research funding through the National Science Foundation (and other agencies), and innovation, measured through renewable energy patents, remains unclear.

The present study leverages federal grants data once housed by the Federal RePORTER database and patenting data provided through USPTO's PatentsView to analyze renewable energy research and patenting trends before and after the deployment of ARRA funding for renewable energy R&D. After preparing and isolating renewable energy research grants and patents, we employed an unsupervised machine learning approach to cluster grants by project duration and total fiscal year funding, and quantified patents by renewable energy subtypes. Findings from the analysis of federal grants do not show an increase in renewable energy research activities after ARRA, but patent data does indicate a 2.5 times increase in renewable energy patents. Because PatentsView does not systematically exclude projects related to Department of Energy (DOE) funding the way that Federal RePORTER data does, this may suggest that ARRA funding for renewable energy R&D—which primarily flowed through DOE—did in fact lead to more renewable energy patenting, but we are simply unable to observe the increase in related research activities due to data limitations with the Federal RePORTER database. It could also indicate that research grants deployed by NSF and other non-DOE agencies yield fewer patents than those produced from DOE-funded projects. On balance, our findings suggest that further investments, such as those stipulated in the *Infrastructure Investment and Jobs Act* of 2021 may lead to further renewable energy innovation, commercialized deployment, and subsequent decreases in carbon emissions.

Introduction

In February of 2022, the Intergovernmental Panel on Climate Change released a report detailing the growing effects of climate change and predictions of fast-approaching environmental calamity. Among the findings, researchers concluded that energy generation diversification and renewable energy resources could limit potential harms from climate change (Pörtner *et al.* 2022). Developing renewable energy resources at scale, however, requires not only investment and infrastructure but also funding in research and development (R&D) for both new technologies and for the advancement of existing technologies.

New investment in renewable energy research and development recently came via the *Infrastructure Investment and Jobs* Act of 2021 (IIJA). While this new funding is encouraging, it is unclear whether R&D investments in renewable energy will yield the type of innovation that would lead to an accelerated and meaningful impact on climate change mitigation. To analyze the impact of renewable energy R&D

investments on innovation, we explored funded research projects and patents related to renewable energy in the years following the 2009 American Recovery and Reinvestment Act (ARRA). We expect the data to show that the $43 billion ARRA investment in "clean energy" projects notably increased research activity in this area, and that innovation, measured with patents, also rose in the subsequent years. In summary, our research intends to explore the extent of the relationship between federally-funded renewable energy R&D and the resulting number of patents and more importantly, the policy related implications of this relationship.

Research Question & Theory of Change

Our primary research question is: What is the extent of the relationship between government-funded research and related patents in the area of renewable energy technology? To be more specific, our research agenda interrogates whether federal funding for basic research in the area of renewable energy leads to renewable energy patents. While the first part aims to determine if there is a link between federal research and development and new climate mitigation technologies, the second part of the analysis will explore the extent of the relationship between grants and patents. The third part would attempt to establish the causal relationship between which grant in the form of expended funding dollars, PI, and renewable type contributed to patent approvals, and ultimately, infer a likelihood of the most effective combination of award types and patents.

ARRA is one of the most recent examples of large investments in renewable energy, so this reinvestment act is a useful policy change to explore, and its time period matches the time period given in our data. Our expectation is to identify an increase in patents as a result of ARRA investments into basic research. A counterfactual we have in mind is had the implementation of the ARRA policy not occurred, there would potentially not have been a dramatic increase in new energy patents. In relation to this logic, a different counterfactual is that in the absence of ARRA, renewable energy research projects and related patents may have seen a substantial decrease, and subsequent decrease in climate innovation.

A theory of change is a description of how a policy intervention might usher in a desired policy outcome. It contains the sequence of events leading to final outcomes (Gertler, 2016).

Our Theory of Change is: Grant funding will lead Principal Investigators (PI's) to invest more time and resources into renewable energy research which will lead to an increase in patent-worthy ideas, systems, and technologies. These "patents"and their associated inducement in technological and commercial innovation might lead to advanced solutions that will address the many facets of modern-era climate challenges.

Our approach in applying an interrogative framework to our theory of change can be summarized as follows:

Activities: Hiring researchers, Collaboration between PIs and Universities or firms, Performance of related research, Drafting and submission of articles for publication, Submission of patent applications. # of PIs.

Outputs: Patents related to different types of renewable energy per year

Outcomes: Provides insight into which type of renewable energy in R&D investment is most investment-worthy.

Final Outcomes: The evaluation of the renewable energy technologies that comprise a large share of a certain type will be patented for mass use. With more people or companies who use these technologies they will be adopted widely and emissions will decline.

In summary, we need to clarify the role of PIs across the research spectrum. Specifically, surface unique findings, innovations, and discoveries that result from funded research activities, published research articles, and patent applications. In terms of outcomes, we expect to be able to identify the types of renewable energy research lines (and PIs) that are most (R&D) investment-worthy. Finally, our hope, and desired policy outcome, is that an increase in renewable technologies will spurr on innovation and commercialization of viable climate solutions by both public and private enterprises.

Literature review

The R&D patent nexus has been broadly researched and narrowly interrogated. At the most general level, research over the years has borne out that there are links between R&D spending and patent activity, as well as in reverse. Danguy, de Rassenfosse, and van Pottelsberghe de la Potterie (2009) re-investigated an earlier study by Almeida and Teixeira (2007) on the relationship between the two by looking across 18 industries in 19 countries. The research revealed that the R&D-patent link was correlated to research productivity and appropriability propensity. Similarly, by looking at 28 EU countries for the period 1999-2013, Sierotowicz (2015) found increases in R&D expenditures (in the business enterprise sectors) had a positive effect on patenting activity. In another study, that took into account not just market considerations but also government interventions in the USA and the EU, Otomo (2017) revealed that while patent counts had increased with expenditures in certain instances, patent applications activity exhibited marginal activity.

Public R&D investments can enhance patenting in the low-carbon energy space. Doblinger's (2019) research is particularly interesting in that it doesn't just link public expenditure in R&D to patents. By looking at clean-tech startups that had been engaged in technology development partnerships with public labs, Doblinger demonstrated a causal link to commercialization. Similarly, Palage (2019) focusing on matching technology-pull policies and public R&D in solar photovoltaic finds similar evidence that public expenditure contributed to increased solar photovoltaic patent activity.

Data & Methods Overview

The data sources we used in our research were CSV files, which were subsets of yearly data that was web-scraped from the Federal Reporter Grants & PatentsView databases. Our original approach started with trying to link together the grant and patents datasets. However, this quickly turned out to be a bad option because some of the patents obtained by the same Principal Investigator were unrelated to the grant's intentions, or vice versa. Instead, we explored and later analyzed both datasets as separate group entities. Through this we could then compare the time periods of ARRA funding and post-ARRA funding and the results. For the grants we filtered on the years 2009-2013, which represented ARRA funding years versus 2014-2018, which represented post-ARRA funding years. Similarly, to account for the lag time of approximately 3 years for patent approval, we looked at patents from 2012-2016, which were the years of ARRA's outcome, versus the years 2017-2018 of Post-ARRA's outcome. The reason we didn't extend our yearly analysis beyond 2018 was because the patent data didn't exist. After an initial exploratory data analysis filtering the grants and patents data by 'renewables' or 'renewable energies' in the titles or abstracts, the most prevalent renewable energy (RE) types were solar, wind, biomass or biofuel, and geothermal. Using these RE types, we can further specify our data to further measure, and visualize the relationship between grants and patents.

Descriptive Statistics

In the grants data from 2009 to 2018, approximately ten thousand rows of data related to renewable energy are from 8 agencies, 50 states, and 1,219 institutions. A total amount of 3.2 billion dollars is invested. Among them, there are 5,430 projects about solar energy, 429 projects about wind energy, 4,236 projects about biomass or biofuel, and 335 projects about geothermal energy. The average funding per project for all renewable energies is roughly 400,000 dollars. More granularly, there is 370,000 dollars on average for solar energies, 290,000 dollars on average for wind energies, 500,000 dollars for biomass, and 250,000 dollars for geothermal. The highest project fund is 146 million dollars from NSF in 2010 for a solar-related project. Even though we knew beforehand that we lacked DOE agency data, the NSF is the agency that invests the most with 7020 projects amassing 2.3 billion dollars in funds. Among those projects, 4,396 projects are about solar energy with 1.7 billion dollars of funding, 385 projects about wind energy with 110 million dollars of funding, 1,932 projects about biomass or biofuel with 770 million dollars of funding, and 307 projects about geothermal energies with 70 million dollars of funding.

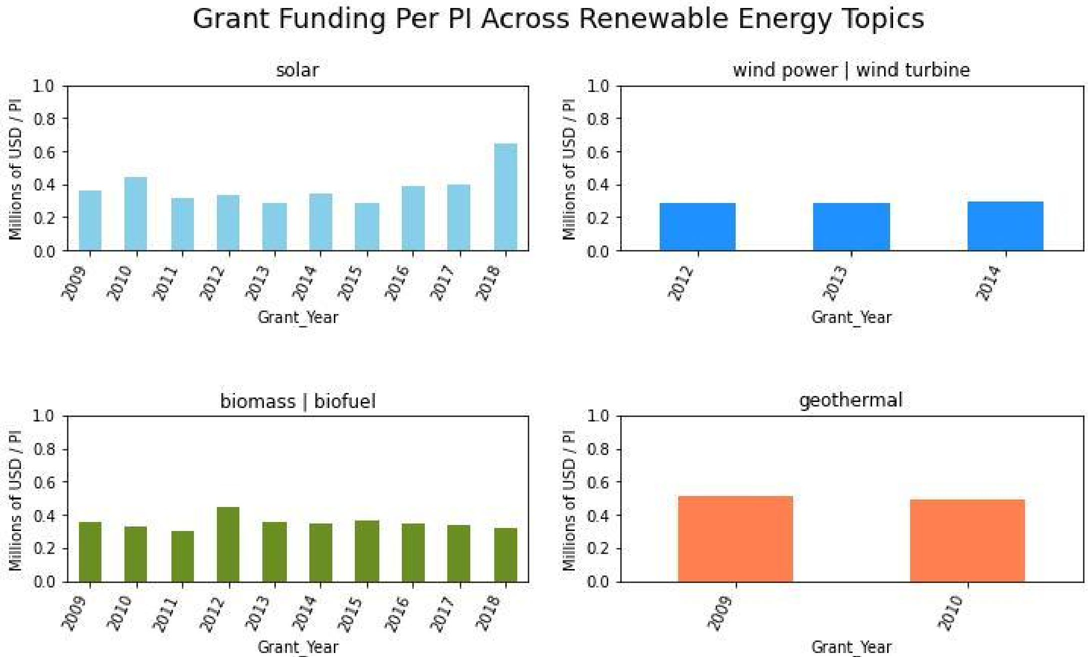
From the data, we can initially see that solar energy accounts for the largest share and receives the largest amount of funds among different types of renewable energy, followed by biomass. Wind and geothermal all have a relatively small proportion. This difference in amount is also reflected in the patent data. From 2009 to 2018, about 6,050 rows of patent data are related to renewable energy from 50 states, 817 cities, and 1,574 institutions. Among them, there are 3,853 patents for solar energy, 890 patents for wind energy, 1,551 patents for biomass or biofuel, and 156 patents for geothermal energy. This comparison between capital and patent generation will provide insights for our process to visualize measures of ARRA's impact.

Grants & Patents Data Pre-processing Methods

Although the impact of ARRA policy involves multiple features that could be measured in different ways, this project mainly focuses on the analysis of the amount of funding. The first key measure we explored in this project is the mean amount of funding per principal investigator.

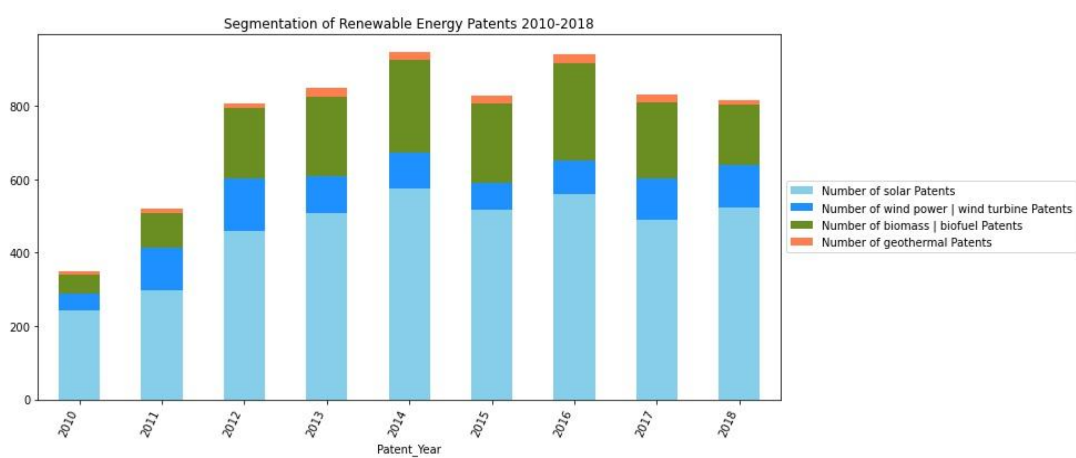
This measure is mainly based on an analysis of grant data, and the data processing is divided into two parts. The first part is data gathering and cleaning. This first step entails gathering data from 2009 to 2018. It is processed by creating a list of the targeted years, then looping through each year, reading in the CSV files each time, and then concatenating the appended list of dataframes versions of the grants or patents. This process is applied separately to both grants and grants abstract data, since they were separated. Through doing this, we can merge them into one dataframe based on the same project ID to obtain abstracts for each row of grants. After that, data cleaning was applied to eliminate rows with no principal investigator. Finally, a column "Grant_Year" is added based on the date the budget started with the project, which serves as easier access to the comparison of data over the years. With this finalized dataframe, we are then able to group the data and plot the amount of fiscal funding per principal investigator each year using the grants data, and the number of patents across years for each renewable energy type.

Visualizations of Grants and Patents Datasets Respectively



Source: FedReporter Grants Database

As shown in this multi barplot, which was included in slide 13 of the presentation, and in the Jupyter notebook titled "Grants-data-processing-EDA-visuals", Solar and Biomass & Biofuel had the most consistent grant funding per principal investigator. We wanted to use this metric to measure the yearly average of funding given to researchers throughout the years. Through using a multi bar plot we would be able to capture any patterns we would have missed if we had just looked at numerical data. Geothermal and Wind REs were mainly funded during the ARRA stimulus. The lack of data beyond those years for those REs shows the impact of missing DOE agency data.

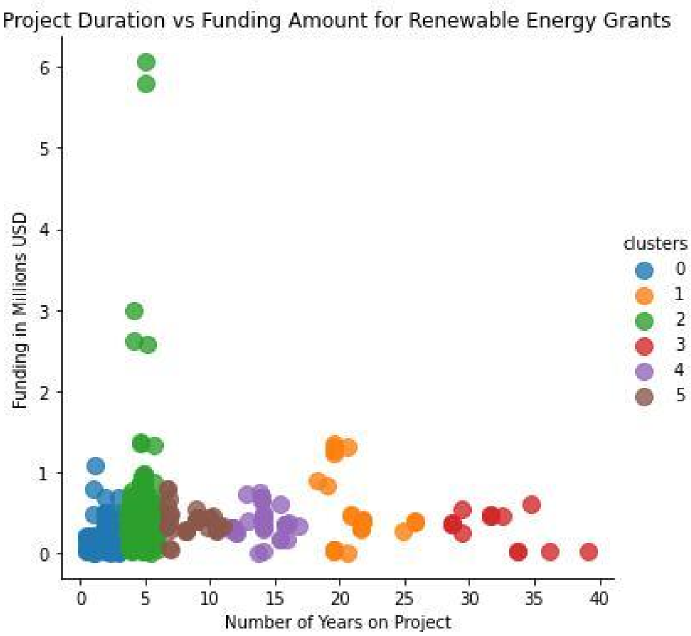


Source: PatentsView Database

Viewing this stacked bar plot, which was on slide 14 of our presentation, and in the Jupyter notebook titled “Patents-data-processing-EDA-visuals”, one can see that renewable energy patents have increased 2.5 times since 2010, which was during the early periods of ARRA funding. It has later approximately plateaued near 800 patents per year post-2014. It can also be seen that solar and biomass or biofuel REs make up the bulk of most patents even after the ARRA stimulus. Most importantly, although the grant data doesn’t contain DOE data, we are able to infer that the patent data does not exclude research funded by the DOE due to the noticeable prevalence of wind and geothermal energies post ARRA.

Unsupervised Machine Learning Approach on the Grants Dataset

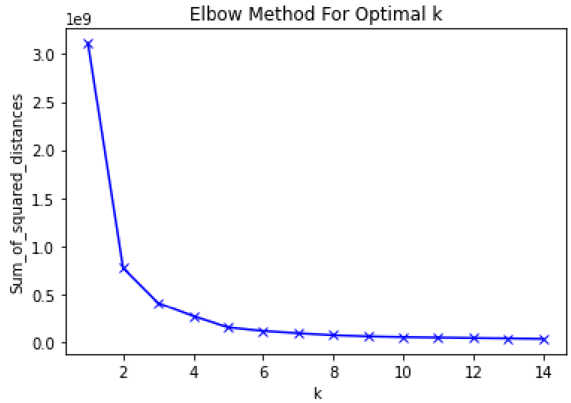
Since we had many renewable grants with few features as well as a lack of renewable energy target labels to fully understand how funding was distributed, it required an unsupervised machine learning approach to the problem rather than a supervised one. A way of implementing this unsupervised method to the data is through the k-means clustering approach. The aim of the k-means clustering algorithm is to partition the n observations in the data into k clusters such that each observation belongs to the cluster with the nearest mean.



Source: FedReporter Grants Database

This scatter plot of the k-means approach on grant project duration and the RE funding amount, which was on slide 12 of our presentation, and in the Jupyter notebook titled “Unsupervised_ML_Grants_data,” shows 6 major clusters. These grant clusters are very similar in certain numerical qualities, where very few exceed past 1.5 million dollars in federal funding. The blue (legend label = 0) and green (legend label = 2) clusters show that a majority of grants have PIs working on them for 2.5 to 5 years, which supports our approach of looking at the grants in ARRA vs Post-ARRA timeframes which last in 4 year periods.

A ‘k’ of 6 clusters was selected based on the elbow optimization method. One would choose a k near the inflection point, where the change in sum of squared distances is negligible or minimized as shown in the image below.



Source: FedReporter Grants Database

Topic modeling for Patent Abstracts

Another approach we wanted to use is text analysis specifically for the patent data filtered on the RE types to clarify understandings about each successfully approved application. Our initial expectations were that there would be more innovative topics generated because of ARRA's impact. After topic modeling, for both the ARRA and Post ARRA time periods, the renewable energy topics were roughly the same across patents. The following three images were obtained from the Jupyter notebook titled "Top_PIs_Text_Analysis", and on slide 15 of our presentation.

After ARRA funding, patent topics:

```
0 plural, plural of, to, assembl, wind
1 biomass, and, of, process, in
2 solar cell, cell, layer, first, second
3 power, system, may, solar, or
```

After Post-ARRA funding, patent topics:

```
0 power, wind, turbin, system, wind turbin
1 biomass, and, to, of, in
2 cell, solar cell, layer, substrat, metal
3 first, second, panel, solar, solar panel
```

Sources: PatentsView Database

Due to this, we wanted to focus more closely on certain PIs that produced the most patents consistently in both time periods. After grouping on the PIs' first and last names to calculate the patent counts for the two time frames of interest, and further inner joining on the same names, we generated the following researchers of interest:

	first_name	last_name	number of patents after ARRA	number of patents after post-ARRA
0	Joseph Broun	Powell	36	17
1	Marshall	Medoff	24	20

Source: PatentsView Database

Looking more closely into the research conducted and the innovation produced by these PIs, biomass was highly prevalent in their patents, so more future federal funding should be assigned to them to create new versions of sustainable fuels or other biofuels. Biomass patents during ARRA perform a more direct operation in converting biomass into fuel, but biomass patents post-ARRA are more likely to aim at providing tunneling techniques, which could be used in ethanol production.

Results, Analysis and Policy Implications

The issue is whether federally funded research into new energy can actually lead to innovation. According to ARRA's policy, renewable energy projects received a significant portion of federal stimulus dollars. As expected, the innovation and progress of new energy projects should follow an influx of funding. From the federal patent data, new energy patents were dominated by solar and biomass projects, followed by wind and geothermal in later years. As a result, since 2010, the renewable energy patents have already increased about 2.5 times, which means ARRA did have a positive impact on the number of new energy patents. However, the nuances of the policy implications here ultimately fall into questions of "viability" and "practicality". These nuances include the types of renewable energy that are most investment-worthy for policy makers to choose in the future. Classifying, matching, and linking the types of funding/agencies to the patent success (in whole and by energy/tech type) can bring greater depth of understanding not just how effective the funding might be, but also how to channel funding when certain energy/tech acceleration is desired.

Solar (Photovoltaic) and Batteries: This technology has come a long way but requires a great deal of land (coverage) and is intermittent and weather dependent. Solar can be paired with battery technologies but they also have scalability and re-use (recycling) issues. In addition, and in conjunction with advanced battery technology, solar technologies rely on Rare Earth Elements which are increasingly difficult to source and are becoming a source of geo-political tension.

Wind: Similar to solar, wind requires large tracts of space, and while they are increasingly being deployed off-shore, they also provide weather dependent and intermittent power.

Both wind and solar are now maturing in terms of technologies but they continue to garner attention due to their low fixed-cost nature - making them economically viable - and their lower cost basis - a direct result of technological and commercial advancements.

Biomass and Geothermal: Both of these technologies have been in use for longer than Solar or Wind, and have come a great deal in terms of being able to be deployed at scale and at greater efficiency. It is not surprising to see our analysis yield no measurable increase given the steady-state nature and longevity of these two energy production methods.

But there are limitations in measuring the direct impact of funding towards innovation in renewable energy patents due to the lack of DOE agency data in the grants data. These problems will be explored more in depth in the next section.

Limitations/biases

The main challenges center on finding connections between funded research projects and subsequent patents, so relied on more basic correlations between annual deployment of research grants and later development of patents. And, while we had a sufficiently large sample size and high-level analysis to link, identify, and quantify the relationships to explore basic machine learning models that might inform future rubric, or criteria, for award selection, there were other important limitations as discussed below.

The first limitation is the uncertainty of the time lag. The length of time it takes for a funded research project in the Federal Reporter grant database to be converted into a patent varies depending on the difficulty of the research. The data we select currently are based on grant data from 2009 to 2018, where it only includes the influence of a 3-year lag, so it may lead to less objective analysis results.

Second, we are unable to link the grant and patent records for useful research. In our analysis, several of the major research project teams with large grant amounts do not appear to be among the teams that received patents. Even in the best of cases, a relatively small share of funded research projects yield a patent. So it is very difficult to link funded research projects with later patent years. The uncertainty around the time lag also increases the difficulty of performing the linkages.

The third issue is the incomplete assessment of the potential value of different types of new energy sources. The output of our data analysis is to provide investors and policy makers with insights into which types of renewable energy are most worthy of R&D investment. However, since the data used only spans a decade, it does not necessarily reflect the most valuable types. Many of the top technology patents have been studied over decades with development and updates, so there are limitations to this data analysis.

The fourth issue is a systematic lack of data. As addressed previously, we do not have grants data from the Department of Energy, which deployed most of the ARRA funds for renewable energy research. If the grants database lacks information on corresponding projects, then some of the patents associated with them in the PatentsView database may be overlooked during data linkages.

Finally, there is also a confidentiality issue that creates a special challenge for our project. Due to national security concerns, much of the DOE data is restricted and confidential. Federal Reporter only has data from NSF and NIH agencies, and we considered whether there is possibly further exploration in the data from DOE. But since they are limited due to the national security issue, we are forced to work with a grants dataset that doesn't contain any.

Conclusion

Overall, after combining the individual studies of renewable energy patent data with the statistical results of federal grant data, we found that, though not stated explicitly in the Patents View data, but through empirical assumptions, DOE funded patents influenced our overall observations of the results of ARRA. After excluding the interference data and missing items, we believe that ARRA has a long-term and positive impact on the innovation and commercialization of new energy patents, as well as the subsequent reduction of carbon emissions and improvement of urban residents' living environment.

Recommendations for future research

There are a number of different directions this project could go in the future. First, the project could benefit from access to Department of Energy grants data. As mentioned earlier, the majority of renewable energy R&D funding in ARRA went to the Department of Energy. Thus, without information directly associated with the DOE, it is difficult to observe the impact of ARRA funds in our grants dataset, which only included a relatively small share that went through, and were masked under the NSF and other federal agencies. If we could access DOE grants data, we could track a much larger share of ARRA-related investments.

A second direction for further investigation may involve the use of record linkage techniques to connect specific grants and patents. For this project we have attempted to identify the "normal" time lag between a grant-funded research project and any subsequent patenting. It would be better if we were able to track specific grant-funded projects to identify which yielded a patent. This isn't the only way to measure innovation related to research investments, but it is a relatively straightforward metric with basis in the literature (Lim *et al.* 2021).

At the same time, we can introduce additional time-series influences. Based on the large amount of grant project data tracked over time, we can analyze approximately how long research projects of different renewable energy types take, classify the difficulty level and include it in our analysis process. A more accurate time lag impact would allow us to more accurately link grant projects to patents in our time series analysis.

Lastly, if we were able to link grants and patents through record linkage techniques, we may then be able to leverage a supervised machine learning approach to predict patent outcomes based on a number of features related to the funded research projects. Such features include whether the PI that was given that grant produced successful patents before or how much funding they received may potentially impact the quality of the patent. Developing a model like this could prove incredibly useful for analyzing the impact of future investments into renewable energy R&D, such as those that have come about through the Infrastructure Investment and Jobs Act of 2021.

References

Alexandre Almeida & Aurora A.C. Teixeira, 2007. "Does Patenting negatively impact on R&D investment?An international panel data assessment," FEP Working Papers 255, Universidade do Porto, Faculdade de Economia do Porto.

Danguy, Jérôme and de Rassenfosse, Gaétan and van Pottelsberghe de la Potterie, Bruno, The R&D-Patent Relationship: An Industry Perspective (December 2010). CEPR Discussion Paper No. DP8145, Available at SSRN: https://ssrn.com/abstract=1718939

Doblinger, Claudia; Surana, Kavita; Diaz Anadon, Laura (2020), "Data for: GOVERNMENTS AS PARTNERS: THE ROLE OF ALLIANCES IN U.S. CLEANTECH STARTUP INNOVATION", Mendeley Data, V1, doi: 10.17632/t8rrxnb6p2.1

Gertler, Paul J., et al. *Impact evaluation in practice*. World Bank Publications, 2016.

Lim, Taekyoung, et al 2021. "The Impact of Intergovernmental Grants on Innovation in Clean Energy and Energy Conservation: Evidence from the American Recovery and Reinvestment Act." *Energy Policy*, Elsevier, 16 Oct. 2020, https://www.sciencedirect.com/science/article/abs/pii/S0301421520306340#fn7.

Otomo, P. (2017). The impact of patents on research & development expenditure as a percentage of gross domestic products: a case in the U.S. and EU economies, Economics Student Theses and Capstone Projects, 55. http://creativematter.skidmore.edu/econ_studt_schol/55

Palage, Kristoffer & Lundmark, Robert & Söderholm, Patrik, 2019. "The impact of pilot and demonstration plants on innovation: The case of advanced biofuel patenting in the European Union," International Journal of Production Economics, Elsevier, vol. 210(C), pages 42-55.

Pörtner, H., et al. 2022. Climate Change 2022: Impacts, Adaptation and Vulnerability, Summary for Policymakers. Working Group II contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change:
https://report.ipcc.ch/ar6wg2/pdf/IPCC_AR6_WGII_SummaryForPolicymakers.pdf.

Sierotowicz, Tomasz. (2015). Patent activity as an effect of the research and development of the business enterprise sectors in the countries of the European Union. JOURNAL OF INTERNATIONAL STUDIES. 8. 101-113. 10.14254/2071-8330.2015/8-2/9.