

U.S. Subsidy Effects on Production of Solar Technology Research: Using a New Measure to Assess Policy Impacts

Eileen Hlavka
Pardee RAND Graduate School
RAND Corporation
Santa Monica, CA, USA
ehlavka@rand.org

Abstract—Solar technology has advanced substantially in recent years, and policymakers interested in reducing climate change hope these improvements will continue. Nevertheless, only one measure, patent counts, has been widely used to identify the rate of improvement, and scant research has assessed the effectiveness of efforts to increase it. In this paper, the number of articles published on solar energy during each month in 1986-2009 is calculated to create a more detailed supplement to patent counts, and these article counts are used to assess certain relevant subsidies. A combination of counting articles manually and naïve Bayesian classification is used to overcome the difficulties of identifying which articles are relevant. The resulting monthly article counts are modeled as a function of two major subsidy types. The largest U.S. subsidy for renewable energy, the Production Tax Credit, appears to be as much as half as effective at encouraging solar technology research as federal solar research subsidies, despite the fact that the tax credit is given to electricity producers rather than to researchers and goes mostly to non-solar energy sources. This finding may be considered a preliminary example of the use of monthly solar article counts.

Solar; renewable energy; technology; learning curve; subsidies; Production Tax Credit; naïve Bayesian classifier; LASSO; text classification; scientometrics

I. INTRODUCTION

Document analysis has rarely been used for understanding renewable energy technological development. Although the current research was inspired in part by Stephens, Wilson et al.'s recent analysis of wind energy news coverage [1] and related work, their research involves no automated document classification and therefore considers fewer articles in greater detail, rather than assessing long-term trends. Across the social sciences more broadly, counting documents by topic is also uncommon.

Prior research on renewable technology development has focused on counting U.S. and European patents. Patents are presumed to reflect valuable innovations, since they are costly to obtain and require demonstration that the concept being patented is new. Patents are also convenient for analysis since they are readily available [see 2, 3], categorized in hierarchical categories, and extend back in time for decades.

Analysis of patent counts for several renewable energy technologies has found that the amount and, in some cases, the type of research responds to energy prices and to subsidies [4-6]. Qualitative and theoretical research suggest similar conclusions [for instance, 7, 8]. Still, the relative sizes of impact of differing subsidy types, such as end-product subsidies vs. research subsidies, have proven difficult to identify [9].

Article counts share many of the benefits of patent counts, although the presumption of value is based on journal editors' decisions. Article counts also have a particular advantage in that they provide finer temporal detail, since approximately as many articles on solar energy are published in a given month as solar energy patents are filed in a given year.¹ However, articles are more complex to collect since they are not well categorized, especially far back in time. This research overcomes that difficulty using a semi-automated method to categorize article abstracts.

The article counts used here are constructed by winnowing over 250,000 abstracts selected by keyword from two academic databases to less than 50,000 relevant abstracts using a combination of hand-coding and naïve Bayesian classification. The abstracts are then sorted by month of publication to create monthly article counts, a new measure of published technology research. Most of the articles included come from engineering or chemistry journals and discuss technical advances in the design, construction and use of photovoltaic panels. The monthly article counts represent the amount of research on solar energy technology published in academic journals in each month during 1980-2009.

While several models are possible for classifying large numbers of documents, a naïve Bayes approach like the one used here is likely to be as efficient and accurate as current competing methods [10]. Naïve Bayes algorithms are fast and predict well while minimizing over-fitting. They have a substantial research history [11] and are well suited for classifying texts such as journal abstracts.

¹ Based on comparison of Popp's patent counts [5] and my article counts for 1986-1993, the only years of overlap.

This research was conducted with the generous support of the Cazier Sustainability Dissertation Fellowship (2009) and a Science to Achieve Results (STAR) Dissertation Fellowship (2010) from the Environmental Protection Agency. I thank these funders, without whom the amount of effort dedicated to this project would have been impossible.

This paper was developed under STAR Fellowship Assistance Agreement no. FP1711801-0 awarded by the U.S. Environmental Protection Agency (EPA). It has not been formally reviewed by EPA. The views expressed in this paper are solely those of Eileen Hlavka, and EPA does not endorse any products or commercial services mentioned in this paper. U.S. Government work not protected by U.S. copyright

Monthly solar article counts are compared with the economic environment in the United States, including federal subsidies, to assess how strongly these subsidies affect solar energy research. The two subsidy types considered are total federal subsidies for solar power research and the Production Tax Credit, which is the largest U.S. federal subsidy for the production of renewable energy. Regression analysis is used to find how many research articles have typically been produced for each dollar of subsidy.

Results suggest that both subsidies have significant effects. In particular, a \$10 million increase in solar technology subsidies leads to a 2% or larger increase in published solar research after one year or more, while a similar increase in the Production Tax Credit leads to a 0.6-1% increase in published solar research after seven months to two years. Given that the tax credit is given mostly to non-solar renewable energy sources and has purposes beyond encouraging innovation, its effect may be considered large. These effects are calculated starting at 2008 levels and larger subsidy increases have correspondingly larger effects.

To the best of the author's knowledge, the present research is the first to use article counts as a way of measuring the rate of research in solar technology and the first to identify the size of effects of the subsidies considered beyond counting articles directly funded by grants. The subsidy results use a simple model and should be considered early results until replicated by additional studies.

II. CONSTRUCTING THE DATA

A. Collecting and Winnowing Articles

This research constructed a new measure of technological developments in solar energy: the number of technical journal articles published on the subject each month. First, potentially relevant abstracts were drawn from two of the most major electronic journal databases. This author and an assistant then sorted a random sample of these abstracts, identifying whether they were relevant or irrelevant. This sample was used to build a naïve Bayesian model of article relevance depending on the words in the abstracts. Each abstract's probability of being relevant was predicted using this model. For each database, after discarding problematic abstracts, the resulting probabilities were counted by month of publication to form a time series representing the approximate number of relevant articles published each month during January 1986 through December 2009.

The two databases that provide the abstracts are ISI Web of Science and Compendex Engineering Village II. Web of Science, owned by Thomson Reuters, includes over 10,000 "high impact journals" covering the sciences, social sciences and humanities. It aims at a large, largely academic audience [12]. Engineering Village II, provided by Elsevier, covers over 5,600 publications on chemical, electrical, mechanical, mining and civil engineering [13], thus focusing on a narrower, more applied audience.

From each database, I extracted all records from the years 1986 through 2009 (as entered by May 2010) with the keyword "solar." Conference proceedings and abstracts that

did not list a month of publication were excluded. This yielded a set of 171,160 abstracts from Web of Science and 102,878 abstracts from Engineering Village.

Within these abstracts, some address solar energy and most do not. Most of the relevant articles discuss the design of photovoltaic cells, including silicon and thin film surfaces. Others describe control systems for large photovoltaic installations, lenses and fluids for use in concentrating solar power, or the amount of sunlight available for solar electricity at a given location. Abstracts were considered relevant if, like these, they seek to improve technology for the provision of electricity using solar power as the direct source. Related topics that do not meet this definition, such as solar-driven heat production, were defined as irrelevant. Most irrelevant articles cover topics that are farther afield, such as astronomy or plant growth.

From over 250,000 abstracts collected, a random sample of 750 was taken from each database. An assistant and I read each of these abstracts and categorized it as relevant or irrelevant. The two readers agreed on the relevance of 91% of the abstracts, for a kappa value of .79, which shows relatively strong agreement.² We discussed and agreed on a final category for every abstract which we had initially classified differently.

The hand-sorted sample was then used to build a naïve Bayesian classifier. This method was chosen in order to sort a large number of abstracts efficiently, use a large number of predictor variables and minimize over-fitting. The functional form used was a logistic regression model with a Laplace prior distribution, solved via maximum a posterior approximation.³ The approach taken also can also be considered a type of least absolute shrinkage and selection operator (LASSO). Words appearing in the hand-sorted abstracts were used as the predictors in the model, after several preprocessing steps.⁴

Ten-fold cross-validation shows that the model predicts an article's relevance correctly about 89% of the time. The model performs well across several diagnostics, shown in the table below. For this purpose, articles are considered classified as relevant by the model if it predicts a probability

² Kappa is calculated as the difference between the actual and randomly expected fraction of the time that the author and the assistant agree, divided by one minus the randomly expected fraction of the time that they agree. The random calculation uses the fraction of abstracts that each coder classified as relevant. Kappa is zero if agreement is purely random and one if agreement is perfect. This measure, although imperfect, is less biased by the relative size of the two categories than the raw percentage agreement and is relatively standard in the hand-classification literature.

³ The Laplace parameter, which controls shrinkage in the LASSO formulation, was chosen to be

$$\lambda = \sqrt{2/J}. \quad (1)$$

where J is the number of unique words that appear in the data.

⁴ The number of appearances of each word was counted after reducing words to their roots and removing everyday common words and words that appeared fewer than fifty times throughout the sample. Cosine normalization was applied to each abstract's word count list in order to adjust for different lengths of abstracts.

TABLE I. CROSS-VALIDATION OF CLASSIFICATION MODEL RESULTS

Statistic	Percent Error	Precision	Recall	ROC
Mean	11%	84%	81%	95%
Variance Across Runs	5%	19%	21%	0.58%

of relevance greater than 0.5. Variances shown are only the variance across the ten cross-validation runs.

The model was used to predict the relevance of each of the 274,038 abstracts collected. These probabilities were summed by month and database to create two time series, one for each database.

B. U.S. Subsidy and Market Data

Data reflecting concurrent market conditions in the United States, including subsidies, were assembled for comparison with solar article counts. All federal subsidies for *developing* solar technology are summed together, as reported to the OECD and IEA by the U.S. government IEA [14]. Also considered is the Production Tax Credit [15], which is the largest federal renewable energy subsidy in the U.S. [16]. It is a tax credit for each unit of electricity *produced* by renewable power plants, including wind and solar. These data are only available annually but are used directly in later regression results.

Monthly data on more general market conditions were collected from the federal Energy Information Administration [17]. While renewable energy prices vary widely and their averages were not readily available, renewable energy consumption is reported. Prices paid by electric power plants for fossil fuels and nationwide electricity prices were used. To control for the broader

TABLE II. SUMMARY OF U.S. SUBSIDY AND MARKET DATA

Variable	Mean	Standard Deviation	Frequency	Source
Solar Technology Subsidies (million \$) ^a	114	33	annual	OECD / IEA
Production Tax Credit (million \$) ^b	250	244	annual	OMB ^c
Renewable Consumption (million BTU)	546	58	monthly	EIA
Electricity Prices (\$/million BTU) ^b	9.01	0.57	monthly	EIA
Fossil Fuel Prices (\$/million BTU) ^b	2.57	0.75	monthly	EIA
GDP (seasonally adjusted million \$) ^b	12,600,000	1,310,000	quarterly	BEA

a. Inflation-adjusted by OECD/IEA using 2009 CPI.

b. Inflation-adjusted by author using mean of quarterly CPI during 2009.

c. Compiled from annual Analytical Perspectives budget documents published by the OMB.

economic environment, gross domestic product from the Bureau of Economic Analysis [18] was also included, with quarterly data interpolated to monthly by calculating geometric means.

All dollars were inflation-adjusted to 2009 dollars using the Consumer Price Index [19].

III. RESULTS

A. Examining Article Counts

The figures below show the full time series of solar energy article counts from Web of Science and from Engineering Village. Although the data are noisy, they show identifiable peaks and valleys over time, suggesting they may reflect meaningful patterns suitable for further analysis.

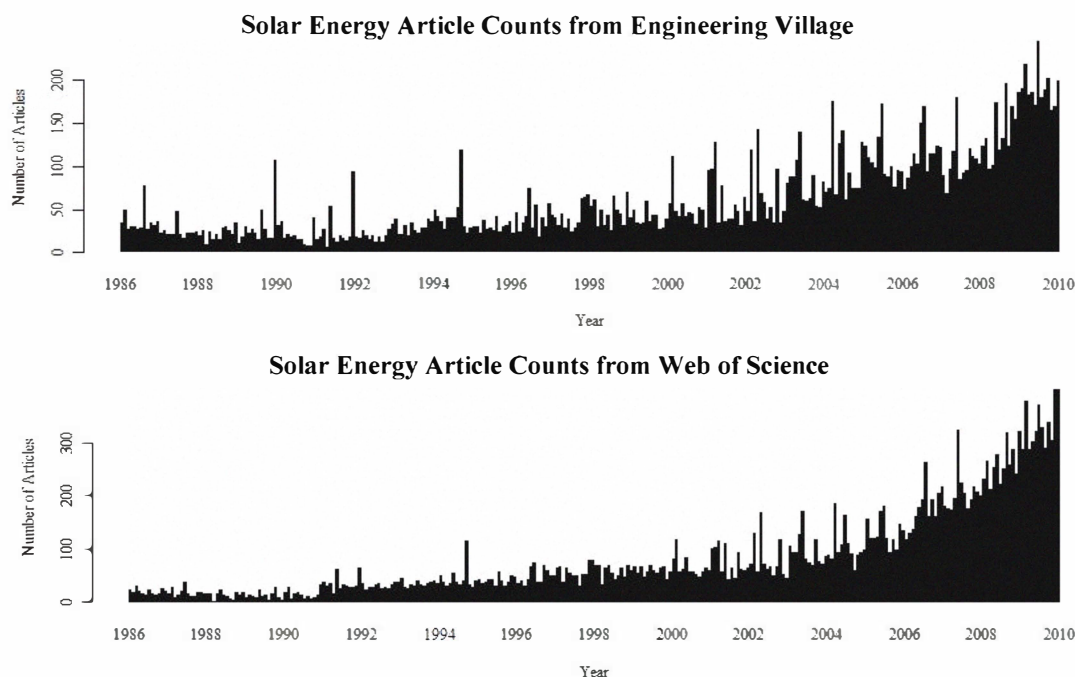


Figure 1. Monthly counts of journal articles on solar energy published in Compendex Engineering Village and ISI Web of Science.

TABLE III. SOLAR ENERGY ARTICLE COUNT SUMMARY STATISTICS

Database	Mean	Median ^a	Standard Deviation	Min.	Max.	Observations
Engineering Village	61	41	48	7	245	288
Web of Science	85	56	84	2	398	288

a. Calculated after rounding each month's count to the nearest integer.

A total of 24,603 relevant articles were identified in Web of Science and 17,550 articles from Engineering Village for 1986-2009. With monthly means of 61 and 85 and no months with zero articles, these data are larger and thus provide finer temporal detail than patent counts. Considering the article counts as a monthly time series, their summary statistics are shown in Table III.

The correlation between Web of Science and Engineering Village monthly article counts is 0.93, with Web of Science counts being higher on average since they come from a larger database.

Several basic patterns appear in the data. As is common for count data, they approximately follow an overdispersed Poisson distribution. An increasing trend is easily visible and autocorrelation and partial autocorrelation plots reveal seasonality as well, with annual and bimonthly cycles presumably arising from the annual schedules of academia and from journals issued bimonthly.

Given this trend and seasonality, one might consider an autoregressive integrated moving average (ARIMA) model. Such a model may be useful for researchers interested in predicting short-term changes in article counts, but it fails to predict major changes in article count slope until after they occur, suggesting that such changes are driven by external factors such as those considered in the next section.

Most of the journal articles included are highly technical discussions of the chemistry and physics of solar cells, especially materials for the photoreactive surface, while a few articles cover other subjects such as usable received sunlight or lenses to concentrate sunlight. As is common for count data, many of the articles are concentrated in a subset of journals, with over one third of articles coming from the top ten journals listed in Table IV, while numerous journals contribute at least half an article—over one thousand journals from each database. Journals that published the most solar energy articles include leading general physics journals and relevant physics and materials science subfield journals as well as solar-specific journals.

B. Modeling Subsidy Effects

1) Regression Model

In order to assess the relationship between subsidies and solar energy article counts, article counts (N_t) were modeled as a function of subsidies and other market conditions. The subsidies considered are total federal subsidies for solar energy research (ss), which captures federal efforts to impact solar research directly, and the Production Tax Credit (PTC), which goes to producers of renewable energy and is the largest federal renewable energy subsidy. Nationwide

TABLE IV. JOURNALS WITH THE MOST SOLAR ENERGY ARTICLES

	Engineering Village		Web of Science	
	Journal	Articles	Journal	Articles
1	Solar Energy Materials and Solar Cells	2,568	Solar Energy Materials and Solar Cells	2,639
2	Thin Solid Films	1,331	Thin Solid Films	1,668
3	Solar Energy	466	Applied Physics Letters	1,225
4	Progress in Photovoltaics	432	Journal of Applied Physics	1,092
5	Solar Cells	381	Renewable Energy	511
6	Renewable Energy	346	Journal of Physical Chemistry C	490
7	Applied Physics Letters	329	Progress in Photovoltaics	453
8	Journal of Non-Crystalline Solids	323	Journal of Physical Chemistry B	442
9	Journal of Crystal Growth	305	Solar Energy	294
10	Journal of Applied Physics	294	Journal of Non-Crystalline Solids	286

a. Some journals have more articles in one database than in the other in part because the databases covered the same journal at different times.

renewable electricity consumption (rc), average electricity prices (ep), and average prices power plants paid for fossil fuels (fp) are also included since they may influence the expected profitability of solar energy innovations. Finally, gross domestic product (GDP) is used to control for the overall state of the economy. Engineering Village and Web of Science article counts are considered separately.

Since article counts are distributed approximately quasi-Poisson, a quasi-Poisson functional form is used to model them, with the canonical logarithmic link function.⁵ All independent predictor variables are lagged by p compared to the dependent variable; that is, article counts at time t are modeled as a function of the values of the other variables p months earlier.

The lag p represents the delay between when an article is written or accepted to a journal, and when it is published. Since it is unclear what value of p best reflects the article creation process, several values are considered. Seven months, the lowest value considered, reflects only the time between article acceptance and publication, as estimated by averaging this time delay across relevant journals, using data collected by Luwel and Moed [20] weighted by the frequency of those journals' appearances in my data. A high of five years is used to represent the longest plausible lag and one or two years is hypothesized to most realistically represent the time between solar research being conducted and being published.

In addition to the predictors described above, a lag of solar energy article counts is included as a predictor variable. This lag captures the effect of existing research at time t on research that is conducted then and published at $t + p$. Theory predicts that the existing "stock" of research affects

⁵ A negative binomial is also appropriate and yields roughly similar results.

the quantity of subsequent research, and inclusion of this stock has been found to affect conclusions about the impacts of other variables such as energy prices [5]. In order to avoid seasonality effects, the lagged version used here is an annual rolling average of monthly article counts up through and including t , written as $ra(N_t)$.

Thus, the full model is

$$\ln(E(N_{t+p})) = \beta_0 + \beta_1 ra(N_t) + \beta_2 ss_t + \beta_3 PTC_t + \quad (2)$$

$$\beta_4 rc_t + \beta_5 ep_t + \beta_6 fp_t + \beta_7 GDP_t$$

with the article counts distributed

$$N_t \sim \text{Poisson}(E(N_{t+p}), \theta E(N_{t+p})). \quad (3)$$

2) Regression Results

These models suggest that both solar technology subsidies and the Production Tax Credit each lead to greater article counts. Solar technology subsidies have the stronger effect, although for Web of Science article counts it is not statistically significant. The tax credit effect is less than half as strong for the same dollar value increase, but this is still substantial given that the tax credit serves other purposes in addition to encouraging solar research. These findings help provide a starting point for quantifying the effects of such subsidies and demonstrate an application of solar article counts.

Because the model is not linear, the coefficients listed in the tables below cannot be interpreted as the effects of unit increases in the predictor variables. For ease of interpretation, the effects of increasing solar research subsidies or the tax credit by fixed percentage and dollar amounts are reported. These are calculated using December 2008 values for all variables as the baseline. Also note that since multiple time lags were not included in a single model, even if a variable has significant effects in multiple time periods, these effects should be interpreted as alternative to each other, not cumulative.

The effect of solar research subsidies on relevant articles is just discernible and is large enough to merit attention. For Engineering Village, a 10% rise in research subsidies yields a significant 4% increase in the number of articles published a year after the funding is disbursed. The same subsidy rise leads to a 9% increase in relevant articles in Web of Science after two years, but this increase is not quite statistically significant. Thus, technology subsidies probably produce a substantial increase in research one or two years after the subsidies are disbursed.

The Production Tax Credit has more obvious effects on research production. A 10% increase (from a much larger base) in dollars disbursed through the tax credit yields a 5%, 8% or 11% increase in articles published in Engineering Village seven months, one year, or two years later. The effect on Web of Science articles is smaller, yielding 1% more articles after one year or 6% more articles after two years. The percent increase in articles from each database scales approximately with the percent increase in subsidies. The average solar energy article published in Engineering

Village may be closer to research application than the average solar energy article in Web of Science, which may explain why Engineering Village article counts are more responsive to production tax incentives.

Given the difference between the current level of spending on solar research and on the Production Tax Credit, it is worth comparing their effects using dollar amounts. Consider an addition of \$10 million. If spent on directly solar technology subsidies, our model predicts an increase of 2% in solar articles published in Engineering Village after one year and perhaps a 5% increase in Web of Science articles after two years.

If the same \$10 million were given out via the Production Tax Credit, it would have a smaller effect, increasing research articles by up to 1%. In Engineering Village, the resulting increase would be 0.6% after seven months, 0.8% after one year, or 1% after two years. For the more basic research in Web of Science, the increase would be 0.2% after one year or 0.7% after two years. Although the tax credit shows a weaker effect on solar research, encouraging solar research is only one among many goals of the tax credit. Thus, this effect may be considered relatively strong.

TABLE V. RESULTS USING SOLAR ARTICLE COUNTS FROM ENGINEERING VILLAGE

Publication Time	7 Months	1 Year	2 Years	3 Years	5 Years
Constant	1.7 (1.4)	2.5 (1.4)	0.072 (2.1)	-0.98 (2.1)	0.93 (2.2)
Lag of Article Counts	0.00037 (0.0051)	-0.0053 (0.0051)	-0.0026 (0.0054)	0.0037 (0.0058)	0.0092 (0.0083)
Solar Technology Subsidies	0.00014 (0.00077)	0.0023 (0.00088)	0.0059 (0.0036)	0.0094 (0.0054)	-0.0033 (0.0059)
Production Tax Credit	0.00057 (0.00022)	0.00082 (0.00037)	0.0011 (0.00039)	-0.00034 (0.00042)	-0.00020 (0.00072)
Renewable Consumption	-0.0014 (0.00062)	0.00015 (0.00062)	0.000218 (0.00072)	0.00063 (0.00072)	0.0015 (0.00082)
Electricity Prices	0.066 (0.066)	-0.22 (0.068)	-0.067 (0.078)	-0.10 (0.093)	-0.067 (0.098)
Fossil Fuel Prices	0.098 (0.095)	0.32 (0.10)	0.053 (0.11)	0.098 (0.12)	-0.14 (0.185)
GDP	$1.8 * 10^{-7}$ ($1.1 * 10^{-7}$)	$2.2 * 10^{-7}$ ($1.1 * 10^{-7}$)	$3.2 * 10^{-7}$ ($1.2 * 10^{-7}$)	$3.7 * 10^{-7}$ ($1.3 * 10^{-7}$)	$3.0 * 10^{-7}$ ($1.3 * 10^{-7}$)
Null Deviance	3,082	2,907	2,585	2,292	1,548
Residual Deviance	1,192	1,112	1,134	1,074	869
Fraction of Deviance Explained	0.61	0.62	0.56	0.53	0.44

a. Coefficients are given, with standard deviations in parentheses. Significance values of $p < 0.5$ are indicated in **bold** while $p < 0.01$ are **in bold and italicized**. Significance is calculated separately for each model.

TABLE VI. RESULTS USING SOLAR ARTICLE COUNTS FROM WEB OF SCIENCE

Publication Time	7 Months	1 Year	2 Years	3 Years	5 Years
Constant	1.5 (1.095)	2.085 (1.250)	0.36 (1.4)	1.3 (1.6)	2.0 (1.9)
Lag of Article Counts	0.0050 (0.0019)	0.0043 (0.0019)	0.0062 (0.0026)	0.0078 (0.0052)	0.029 (0.0082)
Solar Technology Subsidies	-0.0013 (0.00095)	-0.00013 (0.00099)	0.0051 (0.003)	0.00091 (0.0043)	0.0030 (0.0049)
Production Tax Credit	0.00024 (0.00019)	0.00016 (0.00030)	0.00067 (0.00028)	0.00064 (0.00028)	-0.00016 (0.000575)
Renewable Consumption	-0.00081 (0.00045)	0.0011 (0.00046)	0.0013 (0.00052)	0.00095 (0.00053)	0.00046 (0.00065)
Electricity Prices	0.12 (0.054)	-0.13 (0.057)	-0.084 (0.059)	-0.087 (0.067)	-0.056 (0.078)
Fossil Fuel Prices	0.050 (0.058)	0.22 (0.061)	0.10 (0.061)	0.089 (0.0875)	0.068 (0.14)
GDP	$1.5 * 10^{-7}$ $(6.6 * 10^{-8})$	$1.6 * 10^{-7}$ $(7.8 * 10^{-8})$	$2.4 * 10^{-7}$ $(7.9 * 10^{-6})$	$2.2 * 10^{-7}$ $(1.1 * 10^{-7})$	$6.2 * 10^{-6}$ $(1.4 * 10^{-7})$
Null Deviance	5,653	5,430	4,905	4,403	3,259
Residual Deviance	893	862	851	811	837
Fraction of Deviance Explained	0.84	0.84	0.83	0.82	0.74

a. Coefficients are given, with standard deviations in parentheses. Significance values of $p < 0.5$ are indicated in **bold** while $p < 0.01$ are **in bold and italicized**. Significance is calculated separately for each model.

These findings suggest that solar research subsidies and the Production Tax Credit both substantially increase the amount of research published on solar technology. The effect of solar research subsidies is probably stronger but is harder to identify, which may be a result of this subsidy being smaller as well as the noisiness of the article counts. The Production Tax Credit has an effect large enough that it may be a more effective use of funds, depending on what other purposes it accomplishes. While these conclusions should be tested using other models and data sources, they show that solar energy article counts are a meaningful and useful data source.

IV. CONCLUSIONS

This research used a naïve Bayesian model, guided by hand-coded data, to identify journal articles on solar energy. The solar articles were counted and the counts modeled as a function of subsidies and other variables. Results suggest that the Production Tax Credit has a strong and consistent positive effect on the quantity of published solar research, while subsidizing solar research directly has a stronger but harder to ascertain effect. These effect sizes can be subjected to further research and ultimately used as inputs into models predicting the effects of potential subsidy policies. The methods used here could be applied elsewhere to assess other subsidies or types of published research.

ACKNOWLEDGMENTS

In addition to the sponsors, this project has benefited greatly from the influence of Siddhartha Dalal and Nicholas Burger, who provided invaluable guidance throughout. Aimee Curtright gave technical advice on solar energy. Any remaining errors in fact or judgment are my own.

REFERENCES

- [1] J. C. Stephens, G. M. Rand, and L. L. Melnick, "Wind energy in us media: A comparative state-level analysis of a critical climate change mitigation technology," *Environmental Communication: A Journal of Nature and Culture*, vol. 3, pp. 168-190, 2009.
- [2] United States Patent and Trademark Office, Department of Commerce. (2011, May 18). Search for patents. Available: <http://www.uspto.gov/patents/process/search/index.jsp>
- [3] European Patent Office. (2011, May 18). Searching for patents. Available: <http://www.epo.org/searching.html>
- [4] W. A. Pizer and D. Popp, "Endogenizing technological change: Matching empirical evidence to modeling needs," *Energy Economics*, vol. 30, pp. 2754-2770, 2008.
- [5] D. Popp, "Induced innovation and energy prices," *American Economic Review*, vol. 92, pp. 160-180, March 2002.
- [6] D. Popp, "Pollution control innovations and the clean air act of 1990," *Journal of Policy Analysis and Management*, vol. 22, pp. 641-660, Autumn 2003.
- [7] A. Jaffe, Richard Newell and Robert Stavins, "A tale of two market failures: Technology and environmental policy," *Ecological Economics*, vol. 54, pp. 165-174, 2005.
- [8] V. Norberg-Bohm, "Creating incentives for environmentally enhancing technology change: Lessons from 30 years of U.S. Energy technology policy," *Technological Forecasting and Social Change*, vol. 65, pp. 125-148, 2000.
- [9] G. F. Nemet, "Demand-pull, technology-push, and government-led incentives for non-incremental technical change," *Research Policy*, vol. 38, p. 10, 2009.
- [10] J. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naïve Bayes text classifiers," in *Twentieth International Conference on Machine Learning (ICML-2003)*, Washington, D.C., 2003.
- [11] D. Lewis, "Naïve (Bayes) at forty: The independence assumption in information retrieval," in *Machine learning: Ecml-98*, ed: Springer Berlin/Heidelberg, 1998.
- [12] Thomson Reuters. (2010, November 10). Web of science (r) -with conference proceedings. Available: <http://apps.isiknowledge.com/>
- [13] Engineering Village. (2009, November 10). About engineering village. Available: <http://www.engineeringvillage2.org/>
- [14] OECD/IEA, "RD&D budgets, group iii: Renewable energy sources, iii.1 total solar energy," OECD/IEA, Ed., ed, 2010.
- [15] Office of Management and Budget, "Analytical perspectives: Budget of the U.S. Government," for each of Fiscal Years 1996-2011.
- [16] Energy Information Administration, "Federal financial interventions and subsidies in energy markets 2007," April 2008.
- [17] Energy Information Administration, "December 2009 monthly energy review," ed, 2009.
- [18] Bureau of Economic Analysis, "National economic accounts all NIPA tables," ed, 2010.
- [19] Bureau of Labor Statistics, "Consumer price index - all urban consumers," ed, 2011.
- [20] M. Luwel and H. Moed, "Publication delays in the science field and their relationship to the ageing of scientific literature," *Scientometrics*, vol. 41, pp. 29-40, 1998.