Flexibility in the selection of patent counts: Implications for p-hacking and evidence-based policymaking

Peer-reviewed author version

BRUNS, Stephan & Kalthaus, Martin (2020) Flexibility in the selection of patent counts: Implications for p-hacking and evidence-based policymaking. In: Research policy, 49 (10) (Art N°103877).

DOI: 10.1016/j.respol.2019.103877
Handle: http://hdl.handle.net/1942/30322

Flexibility in the selection of patent counts: Implications for p-hacking and evidence-based policymaking

Stephan B. Bruns [a] and Martin Kalthaus [b]

[a] Hasselt University, Centre for Environmental Sciences, Martelarenlaan 42, 3500 Hasselt, Belgium and University of Göttingen, Department of Economics, Humboldtallee 3, 37073 Göttingen, Germany
[b] Friedrich Schiller University Jena, Chair of Economics/Microeconomics, Carl-Zeiß-Straße 3, 07743 Jena, Germany, martin.kalthaus@uni-jena.de

October 20, 2019

Abstract

This study analyzes how researchers' degrees of freedom in selecting patent counts influence econometrically estimated policy effects. Using the evaluation of solar energy policies as an example, we identify 51 strategies to select solar patents from the literature resulting in 306 different solar patent counts, considering six common quality levels of patents. We replicate two leading studies in this literature and re-estimate their econometric models using all of these patent counts. Our results demonstrate severe uncertainty regarding sizes and even signs of key policy effects, opening up the potential for p-hacking and posing a fundamental challenge for evidence-based policymaking. We recommend that more emphasis should be devoted to patent selection procedures, including careful sensitivity analysis regarding key assumptions, such as search strategy and patent quality level. More research is needed to develop common quality standards in working with patent data.

Keywords: Patent data, sensitivity analysis, p-hacking, replication, power, solar energy technologies

JEL Classification: C18, C40, O31

Highlights:

- Researchers' degrees of freedom in selecting patent data are analyzed
- 51 distinct search strategies for solar patents are identified from the literature
- Two leading evaluation studies of renewable energy policies are replicated
- Flexibility in selecting patent counts opens up the potential for p-hacking
- Common standards are needed to pave the way for evidence-based policymaking

1 Introduction

Patent counts have a long tradition in economics as a measure of inventive and innovative activity and are nowadays used to evaluate innovation policies and to elaborate on technological change and economic growth (see Pavitt, 1985; Basberg, 1987; Griliches, 1990; Nagaoka et al., 2010, for surveys of the literature). However, there is little consensus on how relevant patents can be identified and the quality level to be considered. Such researchers' degrees of freedom in data collection can have substantial influence on the empirical analysis and may even be misused to p-hack (Simmons et al., 2011). In this study, we analyze how researchers' degrees of freedom in extracting solar patent counts from patent databases can result in severe uncertainty regarding the sizes and even signs of econometrically estimated policy effects.

Patent counts are frequently used to evaluate policies designed to foster inventive and innovative activities.[1] Based on the induced innovation concept (Hicks, 1932; Acemoglu, 2002), different kinds of policy instruments can be implemented to induce such activities. These instruments can be distinguished in technology push and demand pull instruments (Bush, 1945; Schmookler, 1966; Mowery and Rosenberg, 1979). The effects of these policy instruments on patent counts have been studied, especially for environmentally friendly technologies. Recent contributions are, for example, Aghion et al. (2016) who use patent counts to show that higher tax-inclusive fuel-prices spur inventive activities in the clean auto industry. Calel and Dechezleprêtre (2016) use patent counts to analyze whether the European Union Emissions Trading System induces inventions in low-carbon technologies.

With respect to climate change, solar energy technologies are of particular importance, mirrored by extensive and differentiated policy support that these technologies received to induce inventive and innovative activities. Solar energy technologies can be divided into photovoltaics (PV) and concentrated solar power (CSP). Two seminal contributions in the environmental innovation literature are at the center of our analysis. Johnstone et al. (2010) use patent counts for solar energy technologies (PV and CSP combined) to test how different policy instruments affect inventive output.[2] Peters et al. (2012) test whether domestic or foreign technology push or demand pull policies influence inventive output in PV. Both studies show that direct R&D subsidies as well as demand inducing instruments have a

[1] Inventive and innovative activities describe different parts of the innovation process. While inventive activity refers to the discovery or creation of new knowledge, which results, for example, in a patent, the innovation brings this discovery to the market. In the following we refer to inventive activity if it is related to patented output and to innovation if the process is considered.

[2] While replicating Johnstone et al. (2010), we noticed that a dummy for New Zealand was missing in the analysis, which affects some of the conclusions. We contacted the authors and they published an Erratum, Johnstone et al. (2017). Whenever we refer to Johnstone et al. (2010) in the remainder of this article, we mean the corrected findings.

positive effect on patent counts. However, there is still a debate about the effectiveness of these instruments in the literature and especially among policy makers.[3]

These two and other studies select patent data based on a selection process that is complex and non-standardized. While Griliches (1990) warns about possible pitfalls in using patent data, there is no systematic analysis as to how the researchers' degrees of freedom in the selection of patent counts can influence econometric analyses.[4] The patent selection process can be divided into two parts: the search strategy and the quality level of patents. The search strategy refers to how patents are selected based on their content. The most frequently used search strategies are either based on a classification system or on keywords to search in the patent documents (there are a multitude of other approaches as well). Both classifications and keywords target a specific technology, which is an important part of the search strategy. The quality of patents refers to the technological or economic value of an invention. There are a large number of patents with low novelty contents and the value distribution is highly skewed (e.g. Harhoff et al., 2003). Several approaches are possible to attain a minimum level of patent quality for economic analysis (e.g. Harhoff et al., 2003; Squicciarini et al., 2013). We focus in the following on the most prominent and basic approach that distinguishes patents into six quality levels based on their filing and granting procedure (i.e. priority patents, granted patents, claimed priorities, PCT patents, transnational patents, and triadic patents). As different combinations of search strategies and quality levels of patents may appear plausible in analyzing policy effects, uncertainty in estimated policy effects is likely to be introduced due to these researchers' degrees of freedom.

Researchers' degrees of freedom in patent selection may make it easy to engage in what has recently been coined *p*-hacking (Simonsohn et al., 2014). Hacking the *p*-value refers to uncertainty in model selection, which can be used to select models that 'work', that is, they provide estimates that fulfill the researcher's prior and usually confirm the hypothesis of interest, while those models that did not 'work' remain unreported (Hendry, 1980; Leamer, 1985; Bruns and Ioannidis, 2016). Hacking the *p*-value is eased if flexibility in patent selection approaches transmits to a wide range of estimated policy effects from which statistically significant and hypothesis-confirming estimates can be selected. Presumably, only a small share of researchers are willing to engage in a deliberate search for estimates that 'work'. But researchers face substantial uncertainty as to which patent selection approach is best to proxy inventive activity in a specific technology or industry and, thus, it is natural to

[3] With respect to renewable energy technologies in general, Johnstone et al. (2010) also analyzes other types of renewable energy technologies. Several follow up studies exist that further analyze how policy instruments affect patent counts of renewable energy technologies. For example, DechezleprÃªtre and Glachant (2014) study the effect of domestic and foreign demand pull policies for wind power, Nesta et al. (2014) investigate the effect of polices and competition for renewable energies in general and Costantini et al. (2015b) analyze the effects of technology push and demand pull policies on biofuel patents.

[4] Colombelli et al. (2015) point out a possible bias due to different selection approaches, while de Rassenfosse et al. (2014) show how focusing only on patents from one patent office can bias econometric results.

explore the results for multiple patent selection approaches. It becomes p-hacking if the researcher is influenced by motivated reasoning (Kunda, 1990; Bastardi et al., 2011) and only reports the estimated policy effects that 'work' as (s)he believes them to be the best proxies after seeing the estimates.[5] Nevertheless, irrespective of whether researchers p-hack deliberately or not, published empirical findings become biased and policy effects are likely to become exaggerated, misleading policy makers and resulting in misallocation.

Empirical evidence suggests that such selective reporting is widespread in empirical economics. Brodeur et al. (2016) use more than 50,000 estimates extracted from the *American Economic Review*, *Quarterly Journal Economics* and *Journal of Political Economy* to demonstrate that 10-20% of the marginally significant p-values are inflated, that is, non-significant p-values are inflated to become marginally significant p-values. Based on more than 60,000 estimates from 159 research fields in economics, Ioannidis et al. (2017) show that many studies in economics suffer from low power and that point estimates are exaggerated to obtain statistical significance.

We analyze how researchers' degrees of freedom in the selection of patent counts affect econometrically estimated policy effects by first identifying 51 different search strategies for solar energy technologies that are used in other studies to identify solar energy patents. These different search strategies aim to be generic in capturing the inventive activity in the different solar energy technologies (PV, CSP or both). First, we show how patent counts obtained by these 51 search strategies differ in magnitude, overlap and country coverage, although they intend to measure the same inventive activity. Second, we analyze to what extent the flexibility in the selection of patent counts transmits to variation of econometrically estimated policy effects with direct implications for p-hacking and evidence-based policymaking. To this end, we use as benchmark research designs the two main studies that analyze how policy instruments may affect inventive activity in solar energy technologies: Johnstone et al. (2010) and Peters et al. (2012). In these benchmark studies, patent counts are used as dependent variables and various policy variables are considered as explanatory variables. We estimate the effects of various policy instruments for both benchmark research designs by using 306 alternative dependent variables, that is, patent counts obtained by 51 distinct search strategies with six patent quality levels for each search strategy. We analyze the distribution of each estimated policy effect by means of density plots and vibration plots that relate the size of the estimated coefficient to statistical significance (Patel et al., 2015). Third, we analyze which characteristics of the patent selection approaches determine the sizes of estimated policy effects using meta-regressions (Stanley and Jarrell, 1989).

Our results show that the magnitudes of patent counts obtained by the 51 different search strategies for solar energy patents show severe differences. The overlap of the selected patents across the search strategies varies at the technological level, especially between PV and CSP, while the country coverage is nearly uniform across search strategies. The

[5] A philosophy of science perspective on (deliberate and unintentional) p-hacking is given in Wilholt (2009), Wilholt (2013), citeBiKu2017 and Douglas (2000).

different patent quality levels affect the number of patent counts, but the correlation between patent counts of different quality levels is very high. The sensitivity analysis reveals that the different selection approaches, especially the patent quality levels, result in substantial uncertainty regarding the signs and sizes of estimated policy effects for the two benchmark research designs. We obtain for nearly all policy variables positive as well as negative estimates that are statistically significant. For some of the policy effects the uncertainty is reduced if we consider only the relevant technology of the respective study. In fact, our analysis reveals that the core findings of the two studies that are used as benchmark research designs can be supported in terms of the signs of the estimated policy effects but uncertainty regarding the sizes and levels of statistical significance remain. Moreover, our results reveal that all characteristics of the patent selection approach (technology, search strategy and patent quality level) are key determinants of the sizes of estimated policy effects.

Our results demonstrate that researchers' degrees of freedom in selecting solar patent counts result in substantial uncertainty regarding the sizes and even signs of estimated policy effects. Our analysis does not permit us to identify which selection approach is 'best' in identifying solar patent counts, or which selection approach should be used in general to select patent data. But our findings stress the importance of a sound research design and a careful motivation and documentation of why a specific search strategy and patent quality level is chosen. So far, the research community does not agree on common standards of how to work with patent data. Our findings demonstrate the the need to develop standards for reliable and comparable research. Furthermore, research designs using patent data should provide sensitivity analyses to account for systematic variation in results, especially with respect to patent quality levels. This will help to improve reliability and credibility of econometric evaluations of policy instruments that aim to foster inventive activities and is essential to pave the way for evidence-based policymaking.

The paper proceeds by describing the stylized research process when patent data are used and the two main elements of patent selection approaches. Section 3 describes and analyzes the 51 different selection approaches that we identified from the literature. Section 4 presents our empirical strategy and the results. Section 5 provides a discussion and the last Section concludes with recommendations.

2 Selection of patent counts

2.1 Stylized process of obtaining relevant patents

The beginning of each research project should include a precise research question. Based on the research question, the relevant data to answer this question need to be defined and obtained. Caution in this process is especially relevant for empirical research based on patent data. The stylized process from the research question to a set of patents to answer this research question as well as the flexibility in this process is shown in Figure 1. From

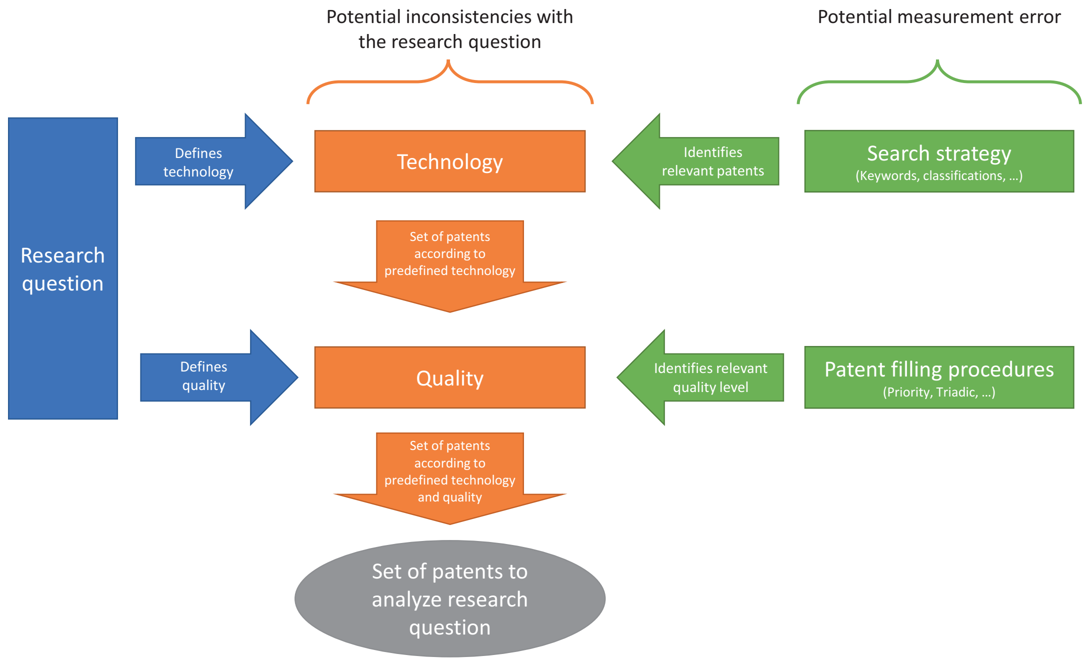

Figure 1: Stylized research process and potential sources of researchers' degrees of freedom.

the research question should follow (i) a clear demarcation of the technology and (ii) a clear definition of the patent quality level. The final set of patents identified for the analysis should be determined (at least) by these two transparently discussed decisions of the researcher.[6]

If there are inconsistencies between the research question and the selection of patents, the research question cannot be answered meaningfully. If, for example, patents are identified based on a similar but different technology as was outlined in the research question or on a narrower or wider interpretation of the scope of this technology, results may become biased in one or the other direction or may even become meaningless. For example, in the case of solar energy technologies, different technologies exist but are often used interchangeably, as discussed in Section 3.

Analogously, the choice of the patent quality level should be guided by the research question. The quality level is frequently used to assess the relevance of inventions and sets the scope of the analysis. Furthermore, it can be used to mitigate potential biases due to different legal settings. For example, in settings where an international comparison is conducted, a quality level that allows for country comparisons should be used to reduce bias towards specific countries (Lanjouw et al., 1998; Dernis et al., 2001; van Pottelsberghe de la Potterie, 2011).

While inconsistencies between the research question, on the one hand, and the analyzed technology or quality level, on the other hand, may easily be spotted by the reader, ambiguities in measurement remain less obvious. Identifying patents that belong to a specific technology requires a search strategy that is able to identify relevant patents. At the same

6 Depending on the research question, further considerations about the patent data may need to be specified, but this is beyond the scope of the present article.

time, patents that may, for example, only share the same technological principle, but do not economically belong to the analyzed technology, need to be excluded. Generally, the different approaches to search for patents may identify too many or too few patents. A clear-cut demarcation of the technology as well as technical understanding by the researcher is required to mitigate possible errors in deriving a patent search strategy. The same holds true for the patent quality level. Several approaches are put forward by the research community to derive different levels of quality, but there is no consensus as to which level should be used for which kind of research question (Dernis and Khan, 2004; Sternitzke, 2009). Of course, each approach has its advantages and disadvantages, which are, however, not always explicit.

The stylized process demonstrates that researchers are left with substantial degrees of freedom in selecting patents. The next two subsections provide an in-depth discussion of the two essential elements of a patent selection approach: search strategy and quality level.

2.2 Search strategies for technology specific patents

The search strategy defines how the relevant patents for a specific technology or product can be selected from patent databases that are managed by various patent offices. Patent offices manage their databases to fulfill their needs – to search for prior art and to clarify the relevance and novelty of the patents they examine. Searching for patents of specific technologies, products, or processes to conduct economic and econometric analyses is possible by several strategies (see Benson and Magee, 2013; Abbas et al., 2014, for overviews). The common strategies to select patents are based on patent classification systems, keywords or their combination (Eisenschitz and Crane, 1986; Dirnberger, 2011; Xie and Miyazaki, 2013). However, each strategy has several advantages and disadvantages that need to be considered when searching for patents.

A patent office uses a classification system that classifies patents according to underlying technological principles to support the examination process (Jaffe and Trajtenberg, 2002). Common classification systems are the International Patent Classification (IPC) managed by the World Intellectual Property Organisation (WIPO) and the recently introduced Cooperative Patent Classification System (CPC). Classification systems are not designed to distinguish specific products or fulfil the economist's needs. This causes problems analyzing specific technologies, products or processes that combine different technological principles. If classification systems are used to search for patents, the respective classifications that describe the specific product or process need to be identified. Here, some issues emerge, since a technological principle can be used for different products or describes different phenomena (Vijvers, 1990; Costantini et al., 2015a).[7] Using classifications might include patents that are not related to the product or process under consideration. This can be referred to as a

[7] For example, the principle of light absorption is characteristic for PV, but digital cameras or other products also refer to this technological principle.

type I error. Additionally, it is also possible that a product or process combines different technological principles and if these classifications are not considered, relevant patents are missing. This can be referred to as a type II error.

Keywords are frequently used to search for patents as well. If a product or technology can be described by a set of technology specific unique keywords, this approach can deliver reliable results. Several issues exist, since keywords can be used in various technologies or products not related to the technology under consideration, resulting in a type I error.[8] Furthermore, a type II error occurs if patent documents are written in a way to avoid specific keywords to keep the invention hidden from competitors. Language differences between patent offices can also reduce the number of selected patents if the keywords do not account for multiple languages or different terminologies used in different countries (Montecchi et al., 2013). In addition, patent databases are not always complete, so that missing titles, abstracts, or other content do not allow searching in all relevant patents. The combination of classifications and keywords can reduce type I errors, but cannot avoid type II errors.

2.3 Patent filing procedures as indicator of patent quality

The second element of patent selection approaches is the quality level. Patents contain codified knowledge, which should be new, non-obvious and applicable. However, the economic and technological value or quality of this knowledge is hard to determine (Griliches, 1990). Several studies estimate the values or qualities of patents (e.g. Schankerman and Pakes, 1986; Harhoff et al., 2003) and find a skewed distribution with a few very valuable patents and many patents with no or little economic value or technological novelty. There are several indicators that can be used to assess patent quality (Squicciarini et al., 2013). One basic indicator is the patent office in which the patent is filed and if it is granted. Since filing in specific or multiple patent offices is expensive, a patent of higher quality or economic value will most likely go through specific filing routes.[9]

The filing procedure can be used as a proxy for patent quality in economic analysis. While for an assessment of inventive activity inside a country, the overall number of patent applications can be sufficient, for international comparisons of inventive activities, patents should be counted on a comparable quality basis because heterogeneity in patent offices can bias results (Lanjouw et al., 1998; Dernis et al., 2001; van Pottelsberghe de la Potterie,

[8] For example, silicon is one of the widely-used materials in PV cells, but it is also used in other semiconductor devices. Furthermore, the product or process can be a non-crucial part of an invention and would be considered as well. In the case of PV, it could be that an electronic device uses a PV cell as power source, even though the invention does not contribute to the PV development at all.

[9] A further frequently used indicator is weighting patents by citations received. However, this procedure opens up further degrees of freedom, since the filing procedure of the patent plays a role, as well as how to count citations.

2011). In the following, we discuss the six different proxies to measure quality levels used in this study and their usefulness for international comparisons of inventive output.

The baseline to count patents is to consider all **priority patents** (first filings) filed at national patent offices. Several shortcomings are attributed to this approach. There are differences between the national patent offices, which can influence the propensity to patent, such as application fees and procedures. Due to this, patents can be quite heterogeneous in terms of quality if different costs to file a patent exist (Atal and Bar, 2010; van Pottelsberghe de la Potterie, 2011; de Rassenfosse and van Pottelsberghe de la Potterie, 2011, 2013).[10]

Granted patents meet some basic quality criteria, since they successfully underwent examination at a national patent office. The criteria to grant a patent, however, differ between countries. This is, for example, reflected in the granting rates (Ordover, 1991), variation in the examination procedures (Lemley and Sampat, 2012), and granting decisions that are not equal across patent offices (Palangkaraya et al., 2011; de Rassenfosse et al., 2016). Eckert and Langinier (2014) discuss further issues related to the granting processes and international comparability.

Claimed priorities refer to patents that have one or several secondary filings at other patent offices claiming a priority filing under the Paris Convention (OECD, 2009). The applicant will only consider an application in a second or more jurisdictions if the patent has the potential to generate revenues to cover these additional costs (OECD, 2009; Haščič and Migotto, 2015). This results in a patent family with members from at least two patent offices. The size of the patent family is a frequently used indicator of patent value (Putnam, 1996; Lanjouw et al., 1998; Harhoff et al., 2003).

PCT patents are applications under the Patent Cooperation Treaty (PCT), which allows the applicant to file a unified application at designated national offices. PCT was introduced in 1970, but it took some time until the procedure was widely accepted; especially in early years only a few applications were filed (OECD, 2009). PCT applications can be of higher value than applications at only the national patent office and allow better international comparability (Grupp and Schmoch, 1999; Guellec and van Pottelsberghe de la Potterie, 2000).

Transnational patents comprise patents either filed via a PCT procedure or filed at the European Patent Office (EPO) (Frietsch and Schmoch, 2010). Both ways of application can impose higher costs on the applicant than national applications and this reduces low-quality applications. They also allow for international comparisons.

Triadic patents are patents with family members filed at the Japan Patent Office (JPO) and the EPO and granted at the United States Patent and Trademark Office (USPTO)

[10] Furthermore, this approach follows the implicit assumption that all the patents that are applied at a certain national office are invented in the respective country. However, this is not always the case and a geographical bias exists, such that especially many applications filed at the United States Patent and Trademark Office are invented in neighboring countries (Griliches, 1990; Harhoff et al., 2009; OECD, 2009). To reduce such bias in international comparisons, patents can be attributed to countries based on the inventor's residence (de Rassenfosse et al., 2013).

(Grupp, 1996; Dernis et al., 2001). These patents are considered of high value due to the imposed costs of filing in these three jurisdictions (Dernis and Khan, 2004; OECD, 2009). They are frequently used for international comparisons, since they have no home country bias (Criscuolo, 2006).

3 Patent counts of solar energy technologies

There is a growing interest in the development of solar energy technologies, especially to mitigate climate change. Numerous studies use patent data to understand how solar energy technologies develop over time, which determinants influence their developments and especially how policy instruments influence the innovation process. Such studies use patent counts or derive further indicators based on patent counts, but they differ in the search strategy used to select the respective patent data.

Solar energy technologies can be divided into two different technologies. Photovoltaics (PV) uses a photovoltaic cell for a direct conversion of sunlight into electricity. Concentrated solar power (CSP) uses a thermodynamic cycle where a collector stores thermal energy in an absorber to utilize the heat for (residential) heating or to use a heat engine to convert heat into electric energy, usually by steam power. Although PV and CSP are often referred to as solar energy technologies, both technologies differ in their applicability, scalability and costs (Peters et al., 2011). The underlying technological principles differ significantly and this results in differences in the innovation processes. The patent search needs to consider these technological differences, especially if patent counts are used for country comparisons, since some countries focus on CSP, others on PV or on both technologies. Technological differences are also relevant for policy evaluations, since countries can implement policies that are not neutral regarding CSP and PV.

We performed a literature search and identified 51 distinct search strategies that are used to select patents of solar energy technologies (see Table 1 and Appendix A).[11] These search strategies differ with respect to the type of search used (classifications, keywords or both) and with respect to the solar energy technology considered (PV, CSP or both). Search strategies intend to be generic for a given technology, as documented by several search strategies being used in multiple studies (see Appendix A). Table 1 presents the search strategies used in the literature disaggregated by type of search and technology.

We use the Worldwide Patent Statistical Database (PATSTAT) (EPO, 2014) to select patent counts based on the 51 search strategies. For each search strategy, we obtain patent

[11] We systematically searched for relevant contributions in the Web of Science, Scopus, Google Scholar, working paper outlets and conference proceedings. We considered only search strategies that used classifications (IPC or CPC), keywords or both. We do not consider strategies using classification or keyword co-occurrences (e.g. Liu et al., 2011; Choe et al., 2013), USPTO classifications (e.g. Popp, 1997; Huang et al., 2011; Benson and Magee, 2013; Dong et al., 2013; Guan and Yan, 2016) or only parts of the technologies (e.g. Tseng et al., 2011; Lizin et al., 2013; Chen and Pham, 2014). In some cases, strategies are not provided (e.g. Watanabe et al., 2001; Marinova and Balaguer, 2009; Breyer et al., 2013).

Table 1: Patent search strategies used in the literature.

	PV	Solar	CSP	Total
Classifications	7	12	6	25
Keywords	11	3	2	16
Both	9	0	1	10
Total	27	15	9	51

Notes: The number of patent search strategies identified from the literature are presented by targeted technology (PV, solar and CSP) and type of search (Classifications, Keywords and both).

counts for the six patent quality levels outlined in Section 2.3 (priority patents, granted patents, claimed priorities, transnational patents, PCT patents as well as triadic patents).[12] We restrict the data to the period 1978-2005 and to 23 countries[13] to be consistent with the two studies that we use as benchmark research designs for our analysis (see Section 4.1).

We present summary statistics for the patent counts selected based on the 51 search strategies by aggregating patent counts across all 23 countries and the period 1978-2005. Table 2 presents these total patent counts for the six quality levels of patents. Total patents vary substantially across search strategies. For example, the maximum number of patents is 243 times larger than the minimum for priority patents. While total patent counts vary across search strategies, relative total patent counts remain similar within each patent quality level as indicated by correlations close to one.

Table 2: Descriptive statistics and correlations.

	No. of patents					Correlations					
	min	median	mean	max	sd	Priority	Granted	Claimed	Transnat.	PCT	Triadic
Priority	2,942	20,946	56,426.0	714,877	135,769.7	1					
Granted	1,714	7,511	20,695.6	266,964	50,625.6	0.9997	1				
Claimed	737	3,762	11,912.6	160,060	30,553.0	0.9991	0.9996	1			
Transnat.	440	2,377	5,837.7	64,619	12,530.9	0.9938	0.9944	0.9943	1		
PCT	290	1,411	3,210.3	34,083	6,629.4	0.9913	0.9926	0.9925	0.9994	1	
Triadic	110	1,109	3,393.6	42,822	8,237.3	0.9967	0.9973	0.9977	0.9983	0.9968	1

Notes: Descriptive statistics for the 51 search strategies identified from the literature are presented by aggregating the patent counts for each search strategy across all 23 countries and the time period 1978-2005. The descriptive statistics are presented for the different patent quality levels. For the aggregated patent counts of the 51 search strategies, correlations between the patent quality levels are shown.

For priority patents, Figure 2 shows the total patent counts for each search strategy. Particularly, search strategies that use classifications can result in extreme patent counts, as

[12] We selected patents by first searching for the respected classifications and/or keywords in all patents in the database and then selected based on the patent's DOCDB Patent Family the priority patent. This approach allows us to capture patents where title or abstract is not available for priority patents. Patents are assigned to a country based on the patent office of the priority patent. Other approaches use the inventor's address, but not all patents contain this information. There are concerns that using the patent office as a proxy for location of inventive activity may introduce bias. As this possible bias would be present across all selection approaches, it is not a relevant concern for our analysis of flexibility.

[13] These countries are: Australia, Austria, Belgium, Canada, Denmark, France, Finland, Germany, Greece, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Republic of Korea, Spain, Sweden, Switzerland, United Kingdom, and United States of America.

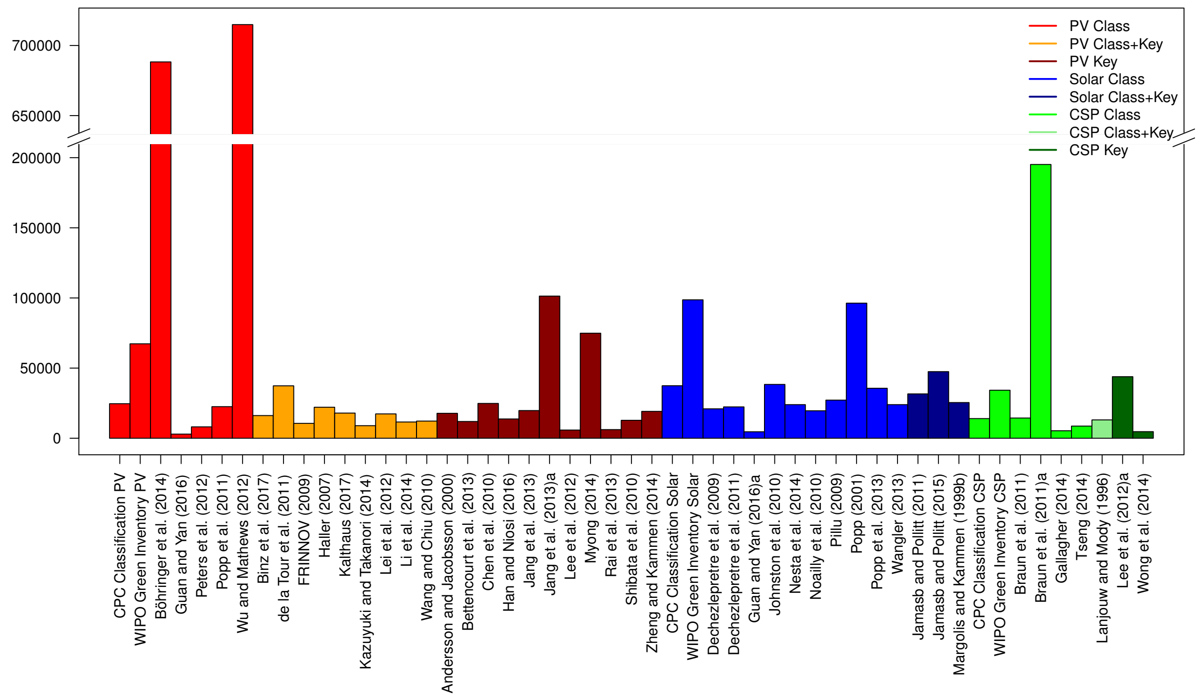

Figure 2: Aggregated patent counts by search strategies (priority patents).

Notes: The horizontal axis shows the 51 search strategies and the vertical axis shows the number of selected priority patents, aggregated for each search strategy across all 23 countries and the time period 1978-2005. 'Class' refers to the use of IPC or CPC classifications and 'Key' refers to the use of keywords.

indicated by both the maximum and minimum of total patent counts stemming from search strategies based on classifications (Wu and Mathews (2012) and Guan and Yan (2016)).[14]

Moreover, the degree of overlap between the sets of identified patents for the various search strategies differ as shown in Figure 3. We measure overlap between two search strategies by the number of patents that are identified by both search strategies and then divide by the number of patents identified by the first search strategy. For example, the overlap of patents between a search strategy based on the CPCs for PV and a search strategy based on the WIPO Green Inventory for PV is 94% with respect to the total of patents in the CPCs for PV and only 34% with respect to the WIPO Green Inventory for PV. This indicates that a search strategy based on the WIPO Green Inventory for PV contains nearly all PV patents, which can be selected by the CPCs for PV, but the WIPO Green Inventory for PV contains a large portion of patents that are not considered in the CPCs for PV.

The geographical distribution of patents identified by the 51 different search strategies is depicted as the share of priority patents per country in Figure 4. Most patents are filed at the Japan Patent Office across all search strategies (more than 65% on average). This high share is related partly to the filing procedure at the Japan Patent Office, which allowed only one claim per patent until 1988 (Sakakibara and Branstetter, 2001). Other countries

[14] See Online Appendix A for descriptive figures for the triadic patent quality level.

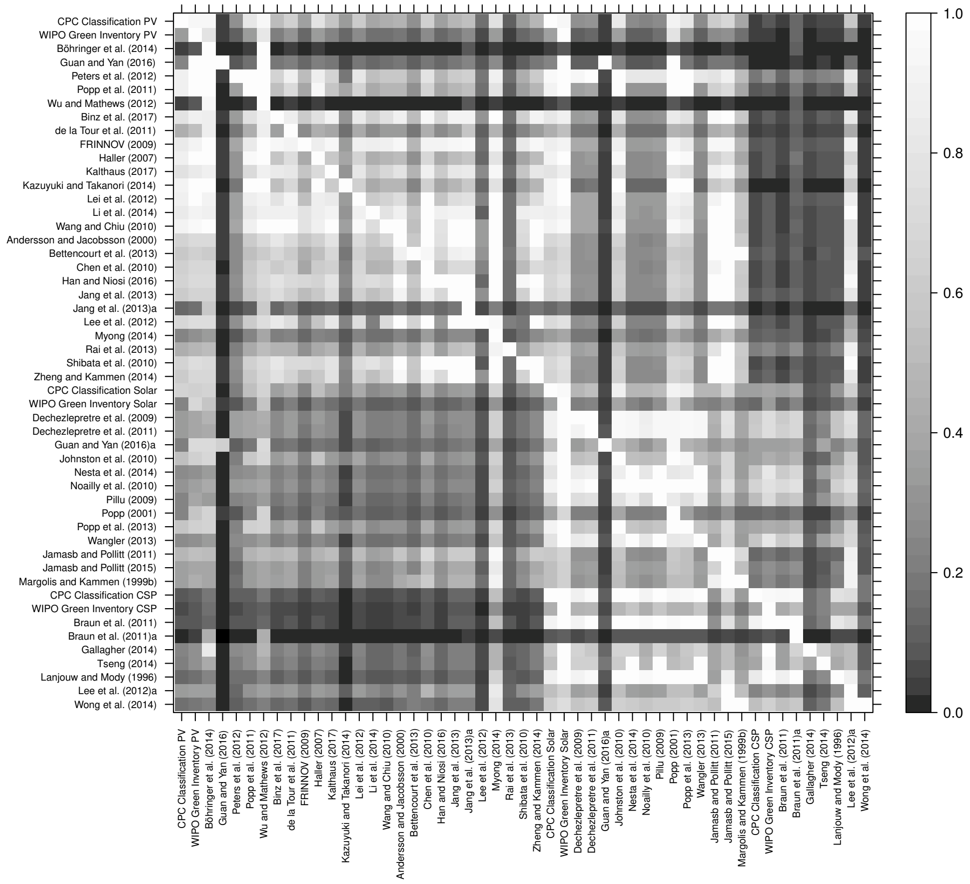

Figure 3: Overlap of priority patents.

Notes: Both axes show the 51 search strategies. Each horizontal line is calculated by $\frac{P_A \cap P_B}{P_A}$ where P_A refers to the number of patents selected by the search strategy on the horizontal axis and P_B to the number of patents selected by the different search strategies on the vertical axis. The lower the overlap of selected patents between two search strategies, the darker the corresponding area.

with a considerable average share are the US (14%), Germany (10%), South Korea (3%), France (3%) and Great Britain (1%).[15]

In a nutshell, total patent counts vary considerably across search strategies, while relative patent counts remain similar across quality levels of patents. The overlap in identified patents varies considerably between technologies. Country shares for selected patents tend to be stable within quality levels, but vary considerably across quality levels.

[15] However, if triadic patents are considered (see Online Appendix A), the US has the highest average share with 37% followed by Japan (27%), Germany (17%), France (6%), Great Britain (3%) and South Korea (2%).

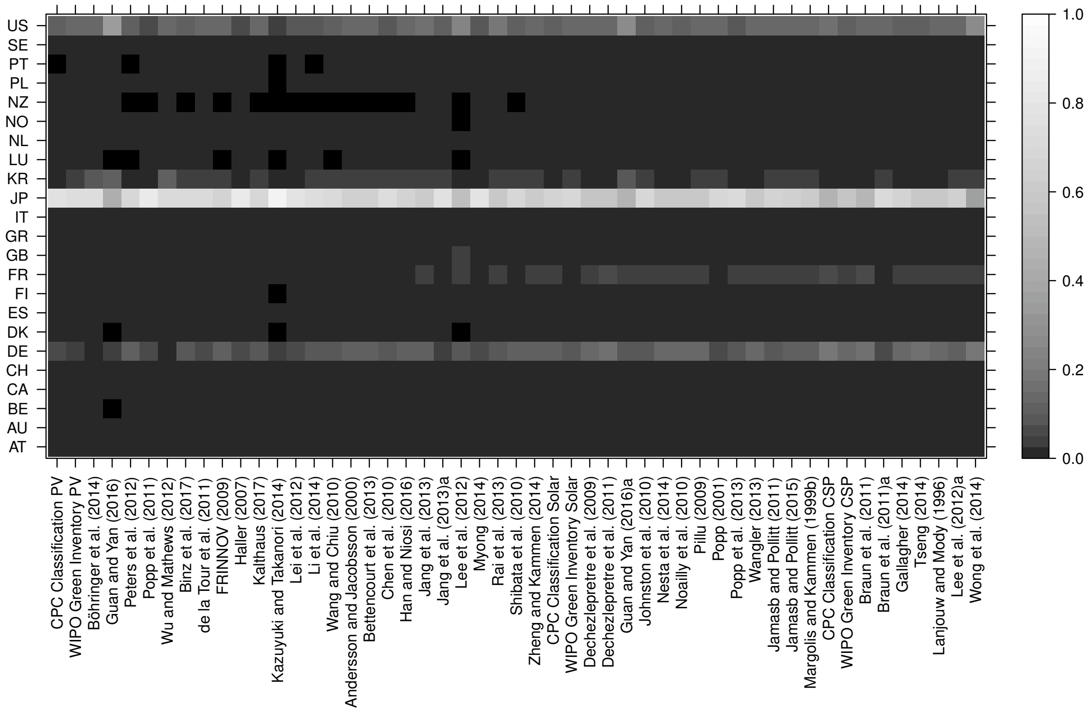

Figure 4: Country shares in priority patents.

Notes: The vertical axis shows the 23 countries in our sample, while the horizontal axis shows the different search strategies. The darker the area, the lower is the country's share in selected priority patents. Black represents no patents at all.

4 Assessing uncertainty in estimated policy effects

4.1 Benchmark research designs

We use the research designs of two leading studies in the field of policy evaluations of renewable energy policies as benchmark research designs for our analysis. These studies are Johnstone et al. (2010) and Peters et al. (2012), which are cited more than 1100 and 270 times respectively, according to Google Scholar (retrieved on October 1, 2019). Both are key contributions to the analysis of demand pull and technology push policy effects on inventive activity. Patent counts are used in both studies to measure inventive output. We use these studies to analyze systematically the variation of estimated policy effects that occurs due to the large variety of patent selection approaches in terms of search strategy, patent quality level and different solar energy technologies.

Both studies use an unbalanced panel of countries and apply a negative binomial model with country fixed effects to estimate how patent counts are influenced by policy and other variables. Johnstone et al. (2010) consider 23 countries for the period 1978-2003 (N=418), while Peters et al. (2012) cover 15 countries for 1978-2005 (N=374). The search strategy of Johnstone et al. (2010) uses several IPCs covering technological principles of PV and CSP. They consistently refer in their analysis to solar energy technologies. Peters et al. (2012) also use IPCs, but they focus explicitly on PV and are restrictive in their search strategy, including only a few classifications. With respect to the patent quality level, Johnstone et al. (2010) use patents filed at the EPO and assign the patents via the inventors' address to countries using fractional counts. Peters et al. (2012) consider three different quality levels, namely triadic patents[16], claimed priorities and all priority patents. For the first two, they assign patents to countries via the inventor's address without fractional counting. For the latter, they use the patent office of the priority patent.

In both studies, several policy variables are considered to analyze their effects on patent counts. Johnstone et al. (2010) use nine different policy variables, three continuous and six dummy variables. The continuous variables are R&D expenditures for solar energy technologies, feed-in tariff levels for solar energy and renewable energy certificate (REC) targets for renewable energy in general. R&D expenditures serve as a technology push instrument, while the other two are demand inducing instruments. The other six variables capture the introduction of several other instruments supporting renewable energies in general, which cannot be measured continuously (Kyoto protocol, investment incentives, tax measures, guaranteed price, voluntary programs and obligations). They find a significant influence of R&D expenditures on inventive activity as well as a weakly significant effect for feed-in tariff levels for solar energy, depending on the model specification. Furthermore, the Kyoto protocol as well as investment incentives have a positive and statistically significant effect on patent output.

Peters et al. (2012) use six variables to capture demand pull and technology push instruments. They use R&D funding for PV to account for technology push. R&D funding is divided into three groups: domestic R&D funding, continental R&D funding and intercontinental R&D funding, to capture domestic and foreign policy effects on domestic inventive activity. Annually installed PV capacity is used to proxy a demand pull effect. Annually installed capacity is again divided into domestic, continental and intercontinental capacity to estimate the effects of domestic and foreign demand pull policies. They find a significant effect for domestic R&D funding for all of their patent quality levels. Domestic and continental capacity shows nearly always a positive significant effect, while a significant effect of intercontinental capacity is only present for priority and claimed patents.

[16] Peters et al. (2012) deviate from the OECD definition of triadic patents and follow Sternitzke (2009). They use "Patent families with publications in at least Germany or the European Patent Office, Japan or China, and the US" (Peters et al., 2012, P 1301, FN 8).

We are able to replicate fully the analysis of Johnstone et al. (2010).[17] For Peters et al. (2012), we can replicate the descriptive statistics of the policy variables but we do not have the patent data to replicate their full analysis.[18]

4.2 Variation of econometric estimates

We assess the variation in estimated policy effects by estimating the regression models of Johnstone et al. (2010) and Peters et al. (2012) for all 306 patent selection approaches (51 search strategies and, for each search strategy, six patent quality levels). The stylized regression to assess the variation of econometrically estimated policy effects is given by

$$Pat_{sq} = P_j \beta_{jsq} + C_j \gamma_{jsq} + \mu_j + \epsilon_{jsq} \tag{1}$$

where $s = 1, ..., 51$ is an index for search strategies and $q = 1, ..., 6$ is an index for patent quality levels. The research design of Johnstone et al. (2010) is denoted by $j = 1$ and the research design by Peters et al. (2012) is denoted by $j = 2$. The research design comprises the study specific policy variables, P_j, control variables, C_j, and country fixed effects, μ_j, including the study specific set of countries and the time horizon considered. The vector of interest is β_{jsq}, which corresponds to the policy variables in P_j. We analyze how $\hat{\beta}_{jsq}$ varies if the dependent variable, the patent counts Pat_{sq}, is varied where $\hat{\beta}_{jsq}$ is the estimate of β_{jsq}.

We use density plots to visualize the uncertainty in estimating $\beta_{jsq}^{(k)}$ due to the researchers' degrees of freedom in selecting patent counts, where k refers to one policy variable in the vector β_{jsq}. We do not expect the density to have a specific shape. Specifically, if measurement error in the dependent variable is random, the estimator remains unbiased but becomes less precise (e.g. Greene, 2012). In this case, we would expect a normal distribution for the estimates $\hat{\beta}_{jsq}^{(k)}$. However, errors in measuring relevant patents corresponding to a specific technology and quality level are likely to be systematic. For example, keywords that cover patents beyond the intended technology are likely to do this differently across countries, e.g. depending on the total number of patents per country. Systematic measurement error results in biased estimation and one may expect to observe basically any distribution for the estimates $\hat{\beta}_{jsq}^{(k)}$. We refrain from linking specific shapes of the densities to selection in the publication process, such as p-hacking, as this would require that all search strategies were designed (at least) for similar research designs compared to the two benchmark research

[17] The estimates in Johnstone et al. (2017) are obtained by using STATA. We can replicate the exact point estimates and model characteristics by using R. However, we cannot replicate the robust standard errors provided by STATA. We use "vcovHC = HC1" from the 'sandwich' package and the standard errors are very similar but tend to be larger. However, these small deviations do not alter the significance levels.

[18] We replicated this study with patent counts based on their search strategy and our point estimates are in the same ballpark. We cannot reproduce the exact patent data of Peters et al. (2012) since they deviate from the OECD triadic patent definition and use a different database (INPAFAMDB - International Patent Family Database) to obtain their patent counts.

designs used in this study, and this is not the case. Moreover, we consider six different quality levels for each search strategy and most of them were not used in the original studies. The focus of this study is to analyze how flexibility in extracting patent counts from patent databases results in uncertainty regarding signs, sizes and levels of statistical significance of estimated policy effects.

Moreover, we visualize the variation in $\hat{\beta}_{jsq}^{(k)}$ by using vibration plots that relate the point estimate of $\beta_{jsq}^{(k)}$ to a transformation of its p-value (Patel et al., 2015; Bruns and Ioannidis, 2016). These vibration plots help to assess the potential for p-hacking by illustrating, for example, whether statistically significant positive and negative estimates can be obtained for the same policy effect.

We focus in our graphical representation on the effects of domestic technology push and demand pull policies as these effects are of particular importance to policy makers. Specifically, we focus on R&D expenditures and feed-in tariff levels from Johnstone et al. (2010) and domestic R&D expenditures and domestic installed capacity from Peters et al. (2012). We show the variation of estimated policy effects for all 306 combinations of search strategies and quality levels as well as for the subset of search strategies designed for the technology that was actually used in the original study. Specifically, for Johnstone et al. (2010) we consider only search strategies for solar, discarding the ones for PV and CSP. In the same manner, we consider only PV search strategies for Peters et al. (2012). This allows us to distinguish between the full uncertainty in estimated policy effects including inconsistencies or imprecisions between the technology outlined in the research question and the technology used to select patents as well as the uncertainty that remains for the particular technology analyzed in the respective study.

As a robustness check, we remove search strategies that result in the 10% largest and 10% smallest sum of priority patents for each technology. As discussed in Section 3, some search strategies identify small sums of priority patents while other search strategies identify huge sums. These extreme search strategies may result in extreme estimates of β_{jsq}. For the robustness check the number of search strategies is reduced from 51 to 39.

4.3 Explaining the variation of econometric estimates

We identify determinants of the variation in $\hat{\beta}_{jsq}^{(k)}$ by using meta-regressions. This approach follows Stanley and Jarrell (1989) who suggest meta-regressions to identify the sources of variation in estimates. While meta-regressions are typically applied to synthesize estimates of multiple studies, we apply meta-regressions to the estimated policy effects obtained by the use of various dependent variables, as outlined in Section 4.1.[19] We estimate

[19] For a similar application of meta-regressions, see Bruns et al. (2018), and for a note of caution regarding the use of meta-regressions to synthesize observational studies to infer the presence of genuine effects, see Bruns (2017).

$$\hat{\beta}_{jsq}^{(k)} = T_s\delta_{1j}^{(k)} + S_s\delta_{2j}^{(k)} + Q_q\delta_{3j}^{(k)} + log(PP_s)\delta_{4j}^{(k)} + \epsilon_{jsq}^{(k)} \tag{2}$$

where $\hat{\beta}_{jsq}^{(k)}$ denotes the point estimate for policy variable k of research design j for search strategy s and patent quality level q with $s = 1, ..., 51$ and $q = 1, ..., 6$ resulting in 306 estimates of $\beta_{jsq}^{(k)}$. As determinants, we consider the analyzed technology, T_s, the type of search strategy, S_s, the quality level of patents, Q_q, and total priority patents of a given search strategy, PP_s. T_s contains two dummy variables. The first is one if PV was used in search strategy s and zero otherwise. The second is one if CSP is used in search strategy s and zero otherwise. S_s also contains two dummy variables. The first is one if keywords are used in search strategy s and zero otherwise, and the second is one if keywords in combination with classifications are used in search strategy s and zero otherwise. Q_s contains five dummy variables that are one if either granted, claimed, PCT, transnational, or triadic patents are used and zero otherwise. Hence, the baseline is the case where the search strategy is based on classifications for the technology solar using priority patents. We also include the log of total priority patents, $log(PP_s)$, of search strategy s to control for the breadth of the search strategy.

4.4 Results

4.4.1 Density and vibration plots

Density and vibration plots for the estimated effects sizes of R&D expenditures from Johnstone et al. (2010) are shown in Figure 5. The first column is based on all 306 patent selection approaches demonstrating the substantial uncertainty in estimating policy effects due to researchers' degrees of freedom in selecting patents. The density plot shows that both positive and negative estimates are possible, with most probability mass for slightly positive effect sizes. The vibration plot shows that statistical significance can be even achieved for both positive and negative effect sizes. The vibration plots show that estimates tend to cluster by the quality level. Specifically, granted patents and priority patents tend to result in positive and statistically significant estimates, while most other quality levels center around zero resulting in a right-skewed density. These results may suggest that R&D subsidies rather increase inventive activity and do not necessarily lead to high quality research output. A possible clustering by search strategy is hard to detect by visual inspection. Only for granted patents does it seem that some search strategies using IPC create larger estimates than keywords and their combination.

The second column of Figure 5 presents density and vibration plots based on search strategies that target solar patents, which is the technology used in Johnstone et al. (2010). This column can be considered as a robustness check of the original study. Negative and significant estimates disappear and the vast majority of estimates are positive, with most

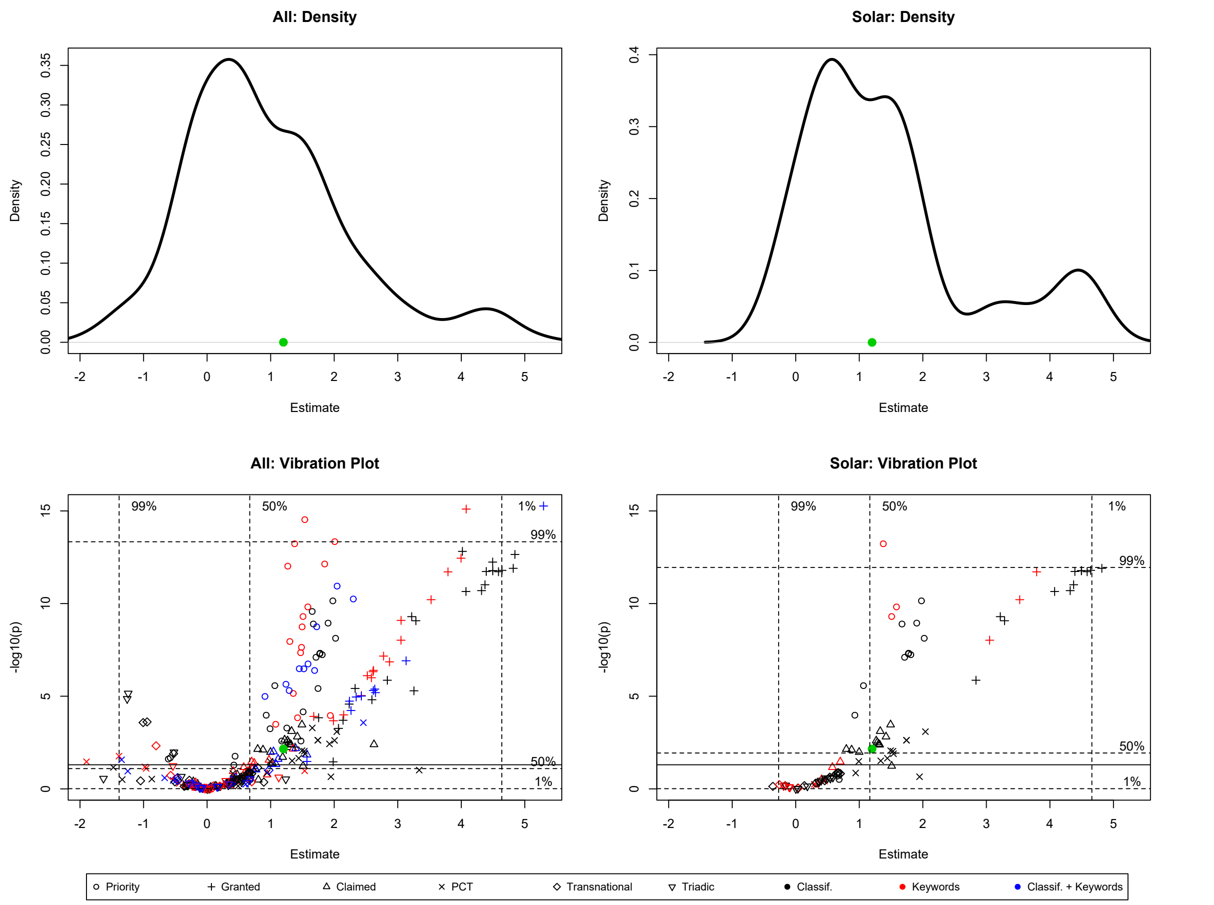

Figure 5: Density and vibration plots for the estimated effect sizes of R&D expenditures from Johnstone et al. (2010).

Notes: In the first row, density plots for the estimated effect sizes are shown. In the second row, vibration plots are presented. In the vibration plot, the y-axis displays $-log10(p\text{-value})$. Thus, all estimates above the solid line are statistically significant at the 0.05 level. "All" covers 306 estimates for all technologies and "Solar" covers 90 estimates for solar technology only. Original estimates are given by green dots as presented in Column 5 of Table 1 (Johnstone et al., 2017).

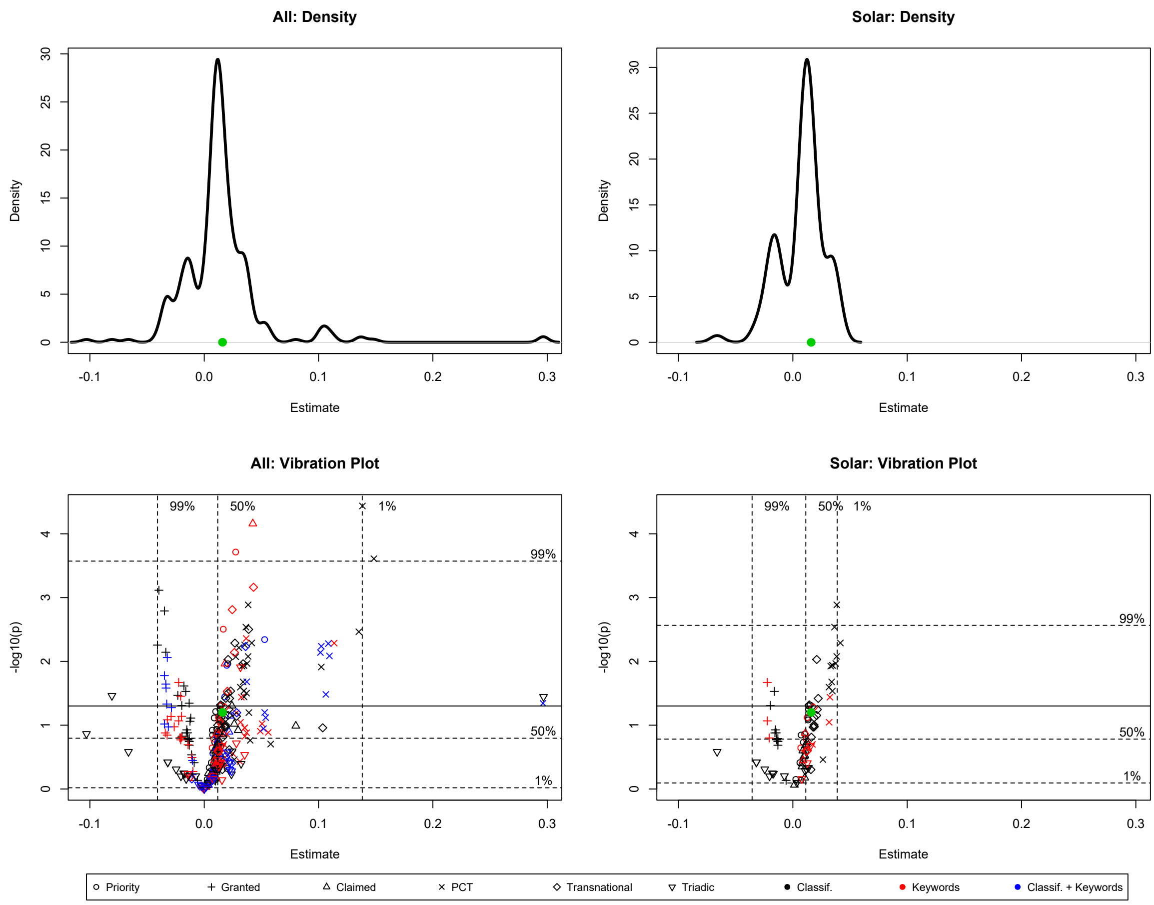

Figure 6: Density and vibration plots for the estimated effect sizes of feed-in tariff levels from Johnstone et al. (2010).

Notes: In the first row, density plots for the estimated effect sizes are shown. In the second row, vibration plots are presented. In the vibration plot, the y-axis displays $-log10(p\text{-value})$. Thus, all estimates above the solid line are statistically significant at the 0.05 level. "All" covers 306 estimates for all technologies and "Solar" covers 90 estimates for solar technology only. Original estimates are given by green dots as presented in Column 5 of Table 1 (Johnstone et al., 2017).

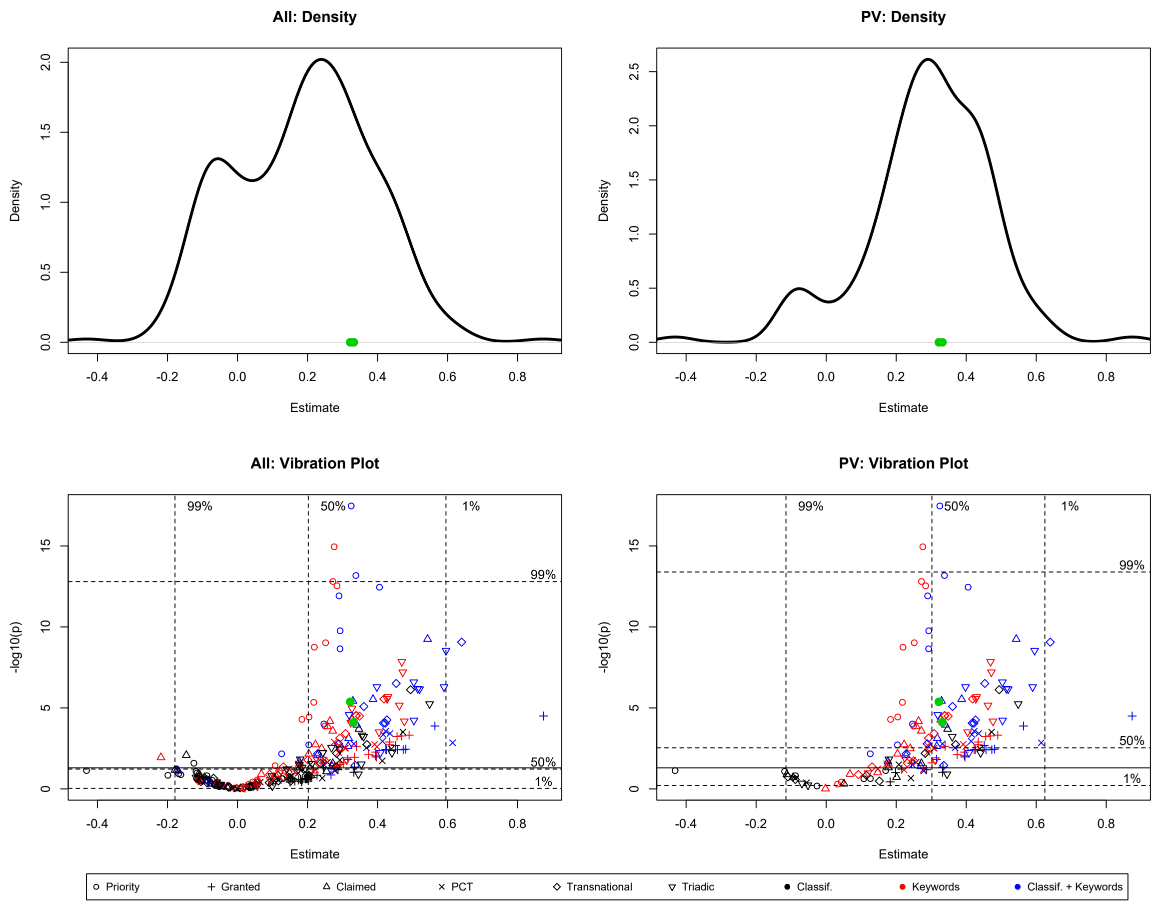

Figure 7: Density and vibration plots for the estimated effect sizes of domestic R&D funding from Peters et al. (2012).

Notes: In the first row, density plots for the estimated effect sizes are shown. In the second row, vibration plots are presented. In the vibration plot, the y-axis displays $-log10(p\text{-value})$. Thus, all estimates above the solid line are statistically significant at the 0.05 level. "All" covers 303 estimates for all technologies and "PV" covers 161 estimates for PV technology only. Original estimates are given by green dots as presented in Column 2 and 4 of Table 2 (Peters et al., 2012).

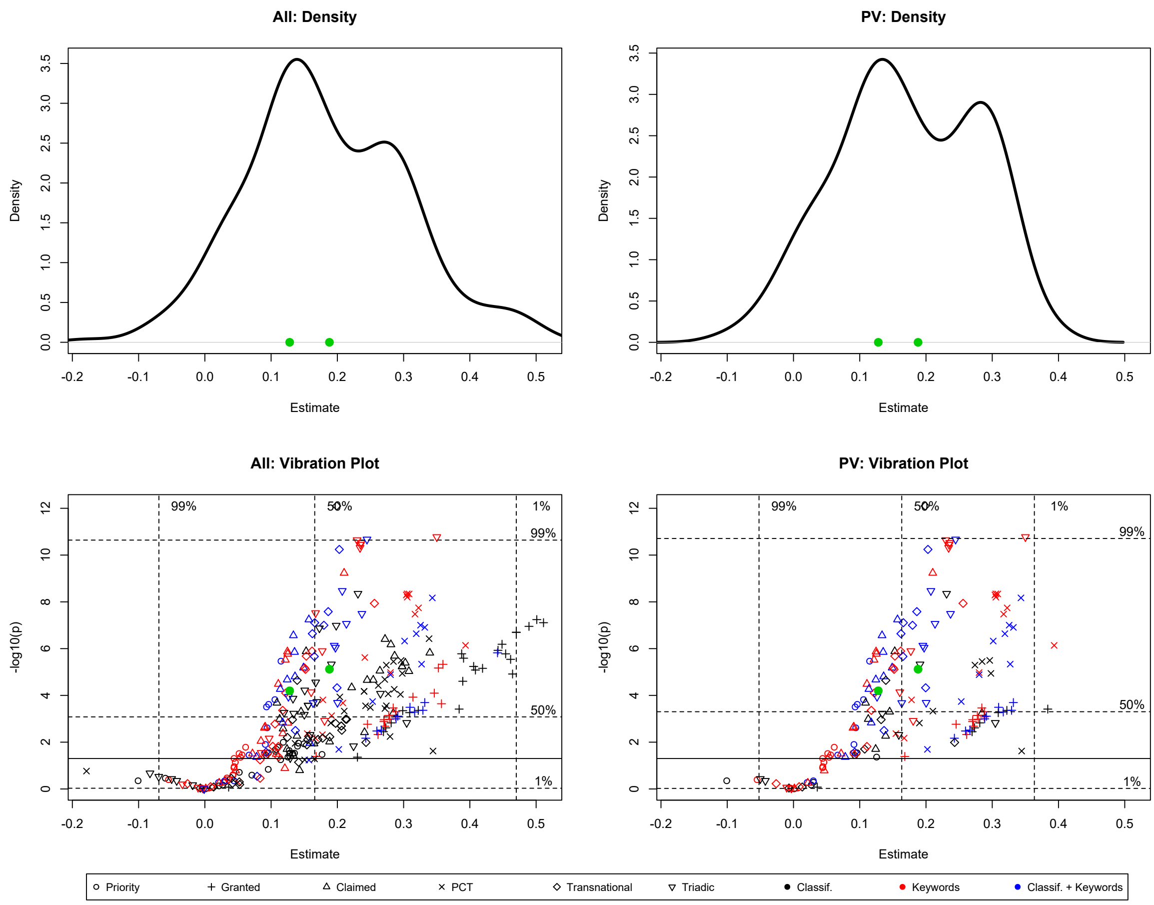

Figure 8: Density and vibration plots for the estimated effect sizes of domestic capacity from Peters et al. (2012).

Notes: In the first row, density plots for the estimated effect sizes are shown. In the second row, vibration plots are presented. In the vibration plot, the y-axis displays $-log10(p\text{-value})$. Thus, all estimates above the solid line are statistically significant at the 0.05 level. "All" covers 303 estimates for all technologies and "PV" covers 161 estimates for PV technology only. Original estimates are given by green dots as presented in Column 2 and 4 of Table 2 (Peters et al., 2012).

probability mass again for slightly positive effect sizes. The density shows a second peak due to the cluster of estimates stemming from the use of granted patents.

Density and vibration plots for the estimated effect sizes of feed-in tariff levels from Johnstone et al. (2010) are given in Figure 6. Statistical significance for both positive and negative estimates can be achieved with most probability mass for slightly positive estimates. Using granted patents results again in a cluster of estimates, but this time with a negative sign. Interestingly, estimated effect sizes for PCT patents cluster here as well, with considerably large and in many cases statistically significant estimates. If only search strategies for solar patents are considered, negative and significant estimates disappear again, except for estimates based on granted patents.

For these two key policy variables, most of the estimates are positive if only search strategies that target solar patents are considered, indicating that the core results of Johnstone et al. (2010) are robust in terms of the signs of the policy effects. This is particularly true when granted patents are ignored. However, uncertainty remains about the effect sizes. The estimates can be interpreted as semi-elasticities and, thus, flexibility in the selection of patent counts results in uncertainty about an increase of solar patents between −0.360% and +4.817% for an additional billion of R&D expenditures. For feed-in tariff levels, the range of effects is −0.066% to +0.042% for an additional US cent/kWh of the feed-in tariff. The extent of variation in econometric estimates is also largely in line with the robustness check that uses a subsample of search strategies by excluding search strategies with extreme patent counts (results are available in the Online Appendix C). Some of the negative and statistically significant estimates disappear.

With respect to the other policy variables used by Johnstone et al. (2010) (see Online Appendix B), similar clustering by quality levels is observable for guaranteed price and voluntary programs in the vibration plots. Interestingly, there are a considerable number of positive significant estimates for higher patent quality levels (PCT, transnational, triadic) for REC targets and investment incentives, but this large number of positive and statistically significant policy effects do not tend to occur if search strategies are restricted to those that target solar patents.

The results for Peters et al. (2012) are given in Figure 7 (domestic R&D funding) and Figure 8 (domestic capacity).[20] The overall findings are similar to those for Johnstone et al. (2010). The clustering of estimates based on granted patents does not seem to be as pronounced as in Johnstone et al. (2010), but clustering by other quality levels can be observed. With respect to the search strategy, systematic patterns are again hard to detect by visual inspection.

If the search strategies for PV are analyzed, which is the technology used in Peters et al. (2012), the estimates for the two key policy variables, namely domestic R&D funding and domestic capacity, are almost exclusively positive indicating that the results of Peters et al.

[20] The estimation procedure in R does not converge for three models for Peters et al. (2012). All of these models use triadic patents. We excluded these three triadic patent counts from the analysis.

(2012) are robust with respect to the signs of the effects as well. Contrary to the results of Johnstone et al. (2010), many significant estimates are based on higher patent quality levels indicating that, for the case of PV, the policy instruments may create higher-valued patents.

Similar to the research design of Johnstone et al. (2010), uncertainty regarding the effect sizes of the key variables remains. Again, trimming for potential outliers in selecting patents does slightly reduce this uncertainty as several significantly negative estimates disappear (results are available in the Online Appendix C). From an economic perspective, the estimates in Peters et al. (2012) can be directly interpreted as elasticities, as the policy variables are considered in logs as well. The effect of increasing domestic R&D funding by 1% on the number of PV patents ranges between -0.431 % and +0.874% and the effect of increasing domestic capacity by 1% between -0.101% and +0.393%.

With respect to the other policy variables used by Peters et al. (2012) (see Online Appendix B), patterns are similar to those of the two main policy variables.

4.4.2 Determinants of estimated effect sizes

Density and vibration plots demonstrate substantial uncertainty regarding signs, sizes and levels of statistical significance of estimated policy effects and suggest that quality levels are a major determinant of effect sizes. Results from the meta-regressions shed further light on the determinants of estimated effect sizes.

For the research design of Johnstone et al. (2010), the results on the determinants of the estimated effect sizes are reported in Table 3. If search strategies for PV rather than for solar are used, the semi-elasticity of R&D expenditures decreases by 0.913 while the semi-elasticity of feed-in tariff levels increases by 0.015. This is well in line with how the policy instruments are used to support the different technologies; R&D expenditures were mostly used to foster research activity for solar technologies in general, while feed-in tariff levels are usually available for PV only. For most other policy variables, using a search strategy for PV increases the semi-elasticities as well. The use of keywords or keywords in combination with classifications does not have an effect on the semi-elasticity of R&D expenditures and only the use of keywords decreases the semi-elasticity of feed-in tariff levels by 0.010. Generally, the use of keywords or the use of keywords in combination with classifications tends to reduce the semi-elasticities in most cases. This is because the classifications either do not capture all the relevant patents or the classifications are too broad, including too many patents. Patent quality levels are decisive for the effect sizes. For R&D expenditures, the semi-elasticity is increased by 1.752 if granted patents are used instead of priority patents. At the same time, using other quality levels, the semi-elasticity decreases, for example, by 1.259 in the case of triadic patents. For feed-in tariff levels, the use of granted patents decreases the semi-elasticity by 0.034, while the use of PCT patents increases the semi-elasticity by 0.045. The highly significant influence of using granted patents on the effect size is also present for all other policy variables and, in some cases, even with a reversed sign compared to the other patent quality levels.

In the case of Peters et al. (2012), results on the determinants of estimated effect sizes are reported in Table 4. Since all variables enter the regression for Peters et al. (2012) in logs, the coefficients can be interpreted as elasticities. Search strategies for PV increase the elasticity for domestic R&D funding by 0.122, while search strategies for CSP decrease the elasticity by 0.127. For domestic capacity, surprisingly, PV search strategies reduce the elasticity by 0.046, while for all other variables, PV search strategies increase the elasticity. The elasticity for domestic R&D funding increases with keywords and keywords in combination with classifications by 0.055 and 0.135, respectively. The effect size of domestic capacity is only affected negatively by search strategies based on keywords. The elasticities of the other variables are, in most cases, also reduced compared to search strategies based on classifications. For domestic R&D funding and domestic capacity, all patent quality levels increase the elasticities. These increases are particularly strong for granted patents, with an increase of 0.261 and 0.269, respectively. For the other variables, there are mixed results, but higher quality levels (PCT, transnational, triadic) usually show positive effects. Again, granted patents show deviating results, especially for intercontinental capacity, where the elasticity is reduced by 0.217 while PCT patents increase this elasticity by 0.090.

Overall, the regression results often have high adjusted R^2 indicating that characteristics of the patent selection approaches are fundamental determinants of the estimated effect sizes. Which characteristic matters most for the estimated effect sizes varies across policy variables. The quality level of patents tends to be important, with granted patents often exhibiting a strong influence on the estimated effect sizes.

5 Discussion

5.1 Uncertainty in estimated policy effects

Our analysis demonstrates substantial uncertainty in econometrically estimated policy effects introduced by researchers' degrees of freedom in extracting patent counts from patent databases. We identify 51 search strategies for solar energy technologies used in the literature and consider six different patent quality levels. The different search strategies lead to severe differences in overall patent counts and the overlap of patents among these search strategies varies considerably. While the country shares in selected do not vary much between search strategies, they differ considerably if different patent quality levels are considered.

Using the research designs of Johnstone et al. (2010) and Peters et al. (2012), we find that researchers' degrees of freedom in the selection of patent counts result in a wide range of estimates for the effects of various policy instruments on patent counts. The uncertainty regarding signs and sizes of these policy effects is substantial, as for almost all policy effects both positive and negative estimates that are statistically significant can be obtained.

Table 3: Determinants of estimated effect sizes for the research design of Johnstone et al. (2010).

	R&D expenditures	Feed-in tariff levels	REC targets	Kyoto protocol	Investment incentives	Tax measures	Guaranteed price	Voluntary programs	Obligations
Intercept	3.808***	0.052**	0.626***	1.699***	-0.383**	-0.454***	1.671***	0.779***	0.895***
	(0.382)	(0.017)	(0.058)	(0.139)	(0.126)	(0.097)	(0.292)	(0.094)	(0.124)
CSP	-0.241*	0.000	0.010	-0.174***	-0.055	0.085**	-0.097	0.008	-0.170***
	(0.113)	(0.005)	(0.017)	(0.041)	(0.037)	(0.029)	(0.086)	(0.028)	(0.037)
PV	-0.913***	0.015**	-0.016	0.002	0.134***	0.172***	0.237**	-0.082***	0.139***
	(0.098)	(0.004)	(0.015)	(0.036)	(0.032)	(0.025)	(0.075)	(0.024)	(0.032)
Class.+Keywords	0.175	-0.005	-0.115***	-0.214***	0.155***	-0.054'	-0.398***	-0.181***	-0.076*
	(0.116)	(0.005)	(0.018)	(0.042)	(0.038)	(0.029)	(0.088)	(0.029)	(0.038)
Keywords	-0.071	-0.010*	-0.102***	-0.250***	0.097**	0.010	-0.309***	-0.181***	-0.134***
	(0.093)	(0.004)	(0.014)	(0.034)	(0.031)	(0.024)	(0.071)	(0.023)	(0.030)
Granted	1.752***	-0.034***	-0.119***	-0.361***	0.348***	-0.328***	-0.479***	-0.893***	0.400***
	(0.127)	(0.006)	(0.019)	(0.046)	(0.042)	(0.032)	(0.097)	(0.032)	(0.041)
Claimed	-0.480***	0.000	-0.030	0.015	0.022	-0.163***	0.103	0.073*	0.213***
	(0.127)	(0.006)	(0.019)	(0.046)	(0.042)	(0.032)	(0.097)	(0.032)	(0.041)
PCT	-0.927***	0.045***	-0.071***	-0.102*	0.768***	-0.197***	0.650***	0.513***	0.531***
	(0.127)	(0.006)	(0.019)	(0.046)	(0.042)	(0.032)	(0.097)	(0.032)	(0.041)
Transnational	-1.279***	0.005	-0.051**	0.072	0.328***	-0.120***	0.204*	0.224***	0.191***
	(0.127)	(0.006)	(0.019)	(0.046)	(0.042)	(0.032)	(0.097)	(0.032)	(0.041)
Triadic	-1.259***	-0.006	-0.086***	0.155***	0.418***	-0.035	0.181'	0.225***	-0.010
	(0.127)	(0.006)	(0.019)	(0.046)	(0.042)	(0.032)	(0.097)	(0.032)	(0.041)
No. of patents	-0.196***	-0.004**	-0.047***	-0.126***	0.017	0.057***	-0.142***	-0.051***	-0.083***
	(0.035)	(0.002)	(0.005)	(0.013)	(0.012)	(0.009)	(0.027)	(0.009)	(0.011)
Adj. R^2	0.757	0.403	0.386	0.484	0.640	0.420	0.369	0.892	0.559
Obs.	306	306	306	306	306	306	306	306	306

Notes: Standard errors in parentheses. Significance levels are depicted by ‘ 0.1, * 0.05, **0.01, ***0.001 level.

Table 4: Determinants of estimated effect sizes for the research desgin of Peters et al. (2012).

	Domestic R&D funding	Continental R&D funding	Intercont. R&D funding	Domestic capacity	Continental capacity	Intercont. capacity
Intercept	0.494***	-0.334***	-0.030	0.309***	0.382***	0.246***
	(0.062)	(0.065)	(0.122)	(0.044)	(0.033)	(0.057)
CSP	-0.127***	-0.093***	-0.094**	-0.087***	-0.037***	0.027
	(0.018)	(0.019)	(0.036)	(0.013)	(0.010)	(0.017)
PV	0.122***	0.167***	0.178***	-0.046***	0.035***	0.044**
	(0.016)	(0.017)	(0.031)	(0.011)	(0.008)	(0.015)
Class.+Keywords	0.135***	0.030	-0.138***	-0.006	0.010	-0.148***
	(0.019)	(0.020)	(0.037)	(0.013)	(0.010)	(0.017)
Keywords	0.055***	0.061***	-0.229***	-0.044***	-0.007	-0.105***
	(0.015)	(0.016)	(0.030)	(0.011)	(0.008)	(0.014)
Granted	0.261***	-0.001	-0.202***	0.269***	-0.039***	-0.217***
	(0.021)	(0.022)	(0.041)	(0.015)	(0.011)	(0.019)
Claimed	0.040’	-0.025	0.127**	0.094***	-0.047***	0.034’
	(0.021)	(0.022)	(0.041)	(0.015)	(0.011)	(0.019)
PCT	0.166***	0.352***	0.322***	0.179***	0.067***	0.090***
	(0.021)	(0.022)	(0.041)	(0.015)	(0.011)	(0.019)
Transtiol	0.127***	0.106***	0.495***	0.066***	-0.045***	0.043*
	(0.021)	(0.022)	(0.041)	(0.015)	(0.011)	(0.019)
Triadic	0.219***	0.156***	0.574***	0.075***	-0.015	-0.040*
	(0.021)	(0.022)	(0.041)	(0.015)	(0.011)	(0.019)
No. of patents	-0.053***	0.011’	-0.014	-0.019***	-0.033***	-0.013*
	(0.006)	(0.006)	(0.011)	(0.004)	(0.003)	(0.005)
Adj. R^2	0.721	0.718	0.674	0.624	0.536	0.575
Obs.	303	303	303	303	303	303

Notes: Standard errors in parentheses. Significance levels are depicted by ‘ 0.1, * 0.05, **0.01, ***0.001 level.

Uncertainty regarding the signs of the core policy variables is substantially reduced if we account for the specific technology used in the original research design. Additionally, excluding search strategies that result in the 10% smallest and 10% largest number of priority fillings helps to slightly reduce uncertainty regarding the signs. However, the economic effect sizes still vary substantially, which makes it difficult to derive recommendations.

For the research design of Johnstone et al. (2010), we can demonstrate that the use of search strategies that focus on the identification of PV patents lead to larger estimates for many policy effects compared to search strategies that aim to identify solar or CSP patents. This indicates that the policy variables considered in Johnstone et al. (2010) seem to encourage inventive activity, particularly in PV and not solar in general, which is the study's original technology choice. This is consistent with the findings for Peters et al. (2012) where estimates of policy effects are in most cases larger if search strategies for PV patents are considered compared to search strategies for solar patents.

With respect to the patent quality levels, we observe deviating results between Johnstone et al. (2010) and Peters et al. (2012). In the former, patent quality levels for higher value patents (PCT, transnational, triadic) result, in many cases, in a decrease in effect sizes, while for Peters et al. (2012), we usually find an increase in effect sizes. Estimates of policy effects for granted patents, however, seem to follow no clear pattern but show usually larger and, in some cases, opposing effects than obtained for the other quality levels. This indicates that granted patents should be used with great care. Many previous studies use granted patents from the U.S. and may need to be interpreted with caution or even reassessed, as most studies present results only for one quality level. As emphasized in Section 2.1, the quality level chosen for the analysis should be transparently derived from the research question and sensitivity analyses should be conducted due to the large influence that quality levels can have on the estimated effect sizes.

We identify characteristics of patent selection approaches as key determinants of policy effects sizes as indicated by large adjusted R^2. All characteristics, namely, search strategy, technology, patent quality level and the total number of identified patents matter in explaining the variation of estimated effect sizes. However, none of the characteristics has a systematic influence on all policy effects besides total priority patents selected by a given search strategy which have mostly a negative influence on the estimated effect sizes. This indicates that the policy effects are targeted towards the technology and do not increase patenting in general, which is likely to be measured if the search strategy is too broad. A precise research question with a clear identification of the relevant technology and an appropriate search strategy can reduce potential measurement errors and reduce potential biases.

Nevertheless, we find that the core results of Johnstone et al. (2010) and Peters et al. (2012) tend to be fairly robust if we consider subsets of search strategies that focus on the respective technologies analyzed in the original studies, even though effects are larger for PV in the case of Johnstone et al. (2010). Especially the two variables measuring technology push and demand pull effects have, in most cases, a high share of positive and statistically significant estimates. This is in line with the theoretical considerations on induced innovation.[21]

5.2 Implications for p-hacking and evidence-based policymaking

Researchers' degrees of freedom in patent selection approaches do not only lead to severe uncertainty with respect to the sizes and levels of statistical significance of estimated policy effects, but also opens up the possibility to deliberately search for results that confirm a hypothesis of interest. The vibration plots intriguingly show that, for almost all policy effects, positive and negative estimates that are statistically significant can be obtained. Moreover, our analysis of the determinants of policy effect sizes reveals that characteristics of the patent selection approaches explain a large fraction of the observed variation. Hence, (small) changes to the selection approach can be used to alter the estimated policy effect until the hypothesis of interest is confirmed. Since patent data for many patent offices and in various forms are nowadays easily available, the deliberate search comes at very low costs.

Although uncertainty opens up the possibility for deliberate p-hacking, we do not believe that empirical research is dominated by researchers who intentionally search for estimates that fulfil their prior beliefs. p-hacking may occur in a much subtler form. Researchers are usually aware of the vast range of estimates that can be often obtained for an effect of interest and they may be affected by motivated reasoning (Kunda, 1990; Bastardi et al., 2011) when they choose to present the empirical model that 'makes sense' according to prior beliefs. There is empirical evidence that such selective reporting is widespread in empirical economics research (Brodeur et al., 2016).

Ioannidis et al. (2017) show that empirical economics research is characterized by low power. This means that the utilized sample sizes are often too small to detect reliably an effect that actually exists. Low power may incentivize authors to p-hack by exaggerating the point estimates in order to find a statistically significant effect. These exaggerations of effect sizes due to p-hacking can be substantial as shown by Ioannidis et al. (2017). Our findings reveal that low power seems to be a problem in the two benchmark studies as well. Some policy effects show a considerably large share of positive but statistically non-significant estimates, indicating that the sample sizes may be too small to ensure standard errors that result in statistically significant estimates. For example, the core variables in Johnstone et al. (2010) (R&D expenditures and feed-in tariff levels) and Peters et al. (2012) (domestic R&D funding and domestic capacity) have, in most cases, a share of positive point estimates close to one. This finding particularly occurred if we use the respective technology of the original research design. Researchers may thus be tempted to use their degrees of freedom in the

[21] As an anonymous reviewer pointed out, the two benchmark research designs do not put much emphasis on the identification of causal effects. In the present study, we can only analyze robustness of the original findings with respect to alternative patent selection approaches. Our analysis does not contribute to the analysis of causality.

selection of patent counts to search for large effect sizes that result in statistically significant estimates. While an explicit discussion of economic significance is desirable (McCloskey and Ziliak, 1996; Cumming, 2014), exaggerated effect sizes may mislead policymakers and result in misallocation. Hence, careful sensitivity analysis regarding the economic significance of estimated policy effects can help to pave the way for evidence-based policymaking.

A rigorous documentation of why a specific search strategy is used and why certain quality levels are considered needs to be implemented to reduce measurement errors and the potential for p-hacking. Particularly, the quality levels of patents need more attention. Peters et al. (2012) can be seen as a good example of how to deal with different patent quality levels, since they demonstrate the flexibility of their results with respect to three patent quality levels. However, such robustness tests towards different measures of patent quality are not common, even in top economics journals.

6 Conclusions and recommendations

We show that the presence of researchers' degrees of freedom in the selection of patent counts calls for a careful interpretation of results obtained with patent data, providing empirical evidence for the warnings made by Griliches (1990). We demonstrate the potential for conscious and unconscious p-hacking by estimating policy effects based on different solar patent counts obtained by different patent selection approaches that vary by search strategy, patent quality level and the type of solar technology considered. Thereby we show how flexibility in the selection of patent counts translates into uncertainty for policy makers in how to evaluate the effectiveness of policy instruments. Moreover, our findings indicate that power may be low in the analysis of patent counts at the country level, potentially incentivizing researchers to exaggerate effect sizes to obtain statistically significant findings (Ioannidis et al., 2017). Exaggerated effect sizes may misinform policymakers about the cost effectiveness of alternative policies and pose a threat to evidence-based policymaking.

We use Johnstone et al. (2010, 2017) and Peters et al. (2012) as benchmark research designs for our comprehensive analysis of variation in estimated policy effects. To this end, our analysis includes replications of these studies with systematic sensitivity analyses. We can show that the key findings of the two studies tend to be robust regarding the signs of the estimated policy effects, which is good news given the large amount of alternative patent counts considered in our analysis. These results are also in line with the theoretical arguments that technology push and demand pull policies induce inventive activities. However, substantial uncertainty remains regarding the effect sizes. For Johnstone et al. (2010), an additional billion of R&D expenditures may result in a change of solar patents between -0.360% and $+4.817\%$ and a one US cent/kWh increase of the feed-in tariff level may result in a change of solar patents in the range of -0.066% to $+0.042\%$. For Peters et al. (2012), a 1% increase of domestic R&D funding may result in a change of PV patents in the range

-0.431% to +0.874% and a 1% increase of domestic capacity may result in a change of PV patents between -0.101% and +0.393%.

Our analyses and findings result in recommendations for future research that involves patent data. We hope that these recommendations may serve as a starting point for a discussion within the scientific community at which end mutually agreed standards for the analysis of patent data emerge.

- More emphasis should be devoted to the patent selection approach as our findings indicate that characteristics of the selection approach can substantially influence econometrically estimated policy effects. For some research designs, Figure 1 in Section 2.1 may prove to be useful to discuss transparently and comprehensively how technology demarcations and patent quality levels directly follow from a precise research question and why they can be best measured by a specific search strategy and quality proxy, including a discussion of potential measurement errors.

- Most importantly, sensitivity analyses regarding the important decisions made in the research process should become obligatory standard. Such a sensitivity analysis imposes relatively low costs on the researcher, since patent databases nowadays provide easy access to the data. Sensitivity analysis seems to be particularly important with respect to the quality levels of the underlying patents, if proxied by the filing procedure, as our findings suggest that the quality level may have a large and heterogeneous influence on econometric estimates. As an example, (Peters et al., 2012) explore robustness of their findings to alternative patent quality levels. Regarding search strategies, new search strategies should be compared to previous ones in the related literature and manual inspection of the obtained patents (ideally by experts) is advised (e.g. Porter et al., 2008; Costantini et al., 2015a). If demarcations based on technological characteristics are needed, narrow and broad search strategies can be compared (e.g. Braun et al., 2011). Insights from future sensitivity analyses may proof useful in developing common standards for research based on patent data.

We have analyzed how flexibility in patent selection approaches transmits to variation in estimated policy effects for a very simple case, the count of patents per country and year. Patent data, however, are used for more sophisticated analyses relying on patents' contents and further meta information. For example, patent data are frequently used to assess technological or economic performance, to map knowledge flows, to reconstruct innovation networks, or to analyze other economic relationships. Such studies are most likely subject to greater flexibility in their results than the two cases presented here, which rely on simple patent counts only.

While we use a specific technology for our analysis, the underlying problem of flexibility is very likely to be present in other technology studies using patent data. For example, similar problems are likely to exist in studies that analyze new and emerging technologies, which are not well captured by specific classifications or keywords such as biotechnology or nanotechnology.

With respect to the researchers' degrees of freedom in patent selection approaches, we focus on three different levels: the search strategy, the patent quality level as well as technological ambiguity. There are further degrees of freedom in the selection and use of patent data that are not considered in the present study. For example, using different classification schemes, selecting patents from different databases or operationalizing patent quality in different ways, such as weighting by patent citations.

Acknowledgments: We are grateful to Nick Johnstone for providing us code and data of both policy variables and patent counts for Johnstone et al. (2010) and to Volker Hoffmann and Michael Peters for providing us data on the policy variables of Peters et al. (2012). This paper was written as part of the research project GRETCHEN (The impact of the German policy mix on technological and structural change in renewable power generation technologies), which is funded by the German Ministry of Education and Research (BMBF) within its funding priority "Economics of Climate Change" under the funding label Econ-C-026. We gratefully acknowledge this support. Previous drafts of the paper were presented at the OECD IP Statistics for Decision Makers Conference 2015 in Vienna, at the Ruhr-University Bochum, the 10th European Meeting on Applied Evolutionary Economics in Strasbourg, the ZEW Mannheim Energy Conference 2017 and the EPIP 2017 conference in Bordeaux. We are grateful for discussions by and with Rudi Bekkers, Uwe Cantner, Holger Graf, Maximilian Göthner, Dietmar Harhoff, Christian Pigorsch, Bastian Rake, Muhammad Faraz Riaz and Karoline Rogge. We are grateful for the comments and suggestions by two anonymous reviewers and the editor.

A Search strategies for solar energy patents

Table 5: Search strategies for solar energy patents

Article	Technology	Type of search	No. of priority patents	Also used in
CPC Classification PV	PV	Class.	24,617	Bointner et al. (2013); Bointner (2014); Stek and van Geenhuizen (2015); Leydesdorff et al. (2015); Wu and Hu (2015); Bäckström et al. (2014); Diederich and Althammer (2016); Breul et al. (2015); Glachant and Dechezleprêtre (2016); Jamali et al. (2016)
WIPO Green Inventory PV	PV	Class.	67,305	Lei et al. (2013); Martinez et al. (2013); Groba and Cao (2015); Choi and Anadn (2014); Gallagher (2014)
Böhringer et al. (2014)	PV	Class.	688,292	Diederich and Althammer (2016)
Guan and Yan (2016)	PV	Class.	2,942	
Peters et al. (2012)	PV	Class.	8,088	Diederich and Althammer (2016)
Popp et al. (2011)	PV	Class.	22,511	Brunel (2016)
Wu and Mathews (2012)	PV	Class.	714,877	Wu (2014)
Binz et al. (2017)	PV	Class. + Key.	16,161	
de la Tour et al. (2011)	PV	Class. + Key.	37,352	
FRINNOV (2009)	PV	Class. + Key.	10,587	
Haller (2007)	PV	Class. + Key.	22,109	
Kalthaus (2017)	PV	Class. + Key.	17,926	Cantner et al. (2016); Herrmann and Töpfer (2017); Kalthaus (2016); Herrmann (2017)
Kazuyuki and Takanori (2014)	PV	Class. + Key.	8,954	
Lei et al. (2012)	PV	Class. + Key.	17,329	Lei et al. (2013)
Li et al. (2014)	PV	Class. + Key.	11,571	
Wang and Chiu (2010)	PV	Class. + Key.	12,228	
Andersson and Jacobsson (2000)	PV	Keywords	17,756	
Bettencourt et al. (2013)	PV	Keywords	11,931	
Chen et al. (2010)	PV	Keywords	24,803	Huang et al. (2013)
Han and Niosi (2016)	PV	Keywords	13,736	
Jang et al. (2013)	PV	Keywords	19,721	
Jang et al. (2013)a	PV	Keywords	101,334	
Lee et al. (2012)	PV	Keywords	5,833	Lee and Lee (2013); Geng and Ji (2016); Wong et al. (2014, 2016)
Myong (2014)	PV	Keywords	74,890	
Rai et al. (2013)	PV	Keywords	6,155	Venugopalan and Rai (2015)
Shibata et al. (2010)	PV	Keywords	12,775	
Zheng and Kammen (2014)	PV	Keywords	19,139	
CPC Classification Solar	Solar	Class.	37,420	Marin and Lotti (2017); Calel and Dechezleprêtre (2016)
WIPO Green Inventory Solar	Solar	Class.	98,668	Albino et al. (2014); Colombelli et al. (2015); Marin and Lotti (2017); Kruse (2016); Kruse and Wetzel (2016)
DechezleprÃªtre et al. (2009)	Solar	Class.	20,946	
DechezleprÃªtre et al. (2011)	Solar	Class.	22,324	Costantini and Mazzanti (2012); Wurlod and Noailly (2016)
Guan and Yan (2016)a	Solar	Class.	4,588	
Johnstone et al. (2010)	Solar	Class.	38,350	Noailly and Shestalova (2017); Noailly and Smeets (2012); Noailly and Ryfisch (2015); Braun et al. (2010); Ayari et al. (2012); Zachmann et al. (2015); Dalmazzone and Corsatea (2012); Diederich and Althammer (2016)
Nesta et al. (2014)	Solar	Class.	23,946	Diederich and Althammer (2016)
Noailly et al. (2010)	Solar	Class.	19557	Noailly (2012); Costantini et al. (2014, 2015c)
Pillu (2009)	Solar	Class.	27,171	Pillu and KolÃ©da (2009)
Popp (2001)	Solar	Class.	96307	
Popp et al. (2013)	Solar	Class.	35,632	Popp (2016)
Wangler (2013)	Solar	Class.	23,946	Corsatea (2014)
Jamasb and Pollitt (2011)	Solar	Keywords	31,601	
Jamasb and Pollitt (2015)	Solar	Keywords	47,483	
Margolis and Kammen (1999b)	Solar	Keywords	25,437	Margolis and Kammen (1999a)
CPC Classification CSP	CSP	Class.	14,004	Wu and Hu (2015)
WIPO Green Inventory CSP	CSP	Class.	34,239	
Braun et al. (2011)	CSP	Class.	14,402	
Braun et al. (2011)a	CSP	Class.	195,172	
Gallagher (2014)	CSP	Class.	5,319	
Tseng (2014)	CSP	Class.	8,657	
Lanjouw and Mody (1996)	CSP	Class. + Key.	13,062	
Lee et al. (2012)a	CSP	Keywords	43,880	Lee and Lee (2013); Geng and Ji (2016)
Wong et al. (2014)	CSP	Keywords	4,693	Wong et al. (2016)

Notes: Search strategies for solar energy patents as used in the literature are presented (collected in March 2017). For each article, the targeted technology, type of search and number of identified priority patents as well as the use of the respective search strategy in other studies is given. 'Class.' refers to the used of classifications (IPC, CPC), 'Key.' refers to the use of keywords and 'Class. + Key.' refers to the use of both. 'PV' refers to photovoltaics and 'CSP' to concentrated solar power and 'solar' to both. Articles marked with an 'a' contain more than one search strategy.

References

Abbas, A., Zhang, L., and Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37:3–13.

Acemoglu, D. (2002). Directed technical change. *Review of Economic Studies*, 69(4):781–809.

Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R., and Reenen, J. V. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, 124(1):1–51.

Albino, V., Ardito, L., Dangelico, R. M., and Petruzzelli, A. M. (2014). Understanding the development trends of low-carbon energy technologies: A patent analysis. *Applied Energy*, 135:836–854.

Andersson, B. A. and Jacobsson, S. (2000). Monitoring and assessing technology choice: The case of solar cells. *Energy Policy*, 28(14):1037–1049.

Atal, V. and Bar, T. (2010). Prior art: To search or not to search. *International Journal of Industrial Organization*, 28(5):507–521.

Ayari, N., Blazsek, S., and Mendi, P. (2012). Renewable energy innovations in Europe: A dynamic panel data approach. *Applied Economics*, 44(24):3135–3147.

Basberg, B. L. (1987). Patents and the measurement of technological change: A survey of the literature. *Research Policy*, 16(2-4):131–141.

Bastardi, A., Uhlmann, E. L., and Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science*, 22(6):731–732.

Benson, C. L. and Magee, C. L. (2013). A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. *Scientometrics*, 96(1):69–82.

Bettencourt, L. M. A., Trancik, J. E., and Kaur, J. (2013). Determinants of the pace of global innovation in energy technologies. *PLoS ONE*, 8(10):e67864.

Binz, C., Tang, T., and Huenteler, J. (2017). Spatial lifecycles of cleantech industries – the global development history of solar photovoltaics. *Energy Policy*, 101:386–402.

Bointner, R. (2014). Innovation in the energy sector: Lessons learnt from r&d expenditures and patents in selected IEA countries. *Energy Policy*, 73(0):733 – 747.

Bointner, R., Biermayr, P., Goers, S., Streit-Maier, J., and Tichler, R. (2013). Wirtschaftskraft erneuerbarer energien Ãsterreich und erneuerbare energien zahlen. Technical report, Blue Globe Report 1/2013, Klima- und Energiefonds.

Braun, F. G., Hooper, E., Wand, R., and Zloczysti, P. (2011). Holding a candle to innovation in concentrating solar power technologies: A study drawing on patent data. *Energy Policy*, 39(5):2441 – 2456.

Braun, F. G., Schmidt-Ehmcke, J., and Zloczysti, P. (2010). Innovative activity in wind and solar technology: Empirical evidence on knowledge spillovers using patent data. Technical Report 7865, CEPR Discussion Papers.

Breul, M., Broekel, T., and Brachert, M. (2015). Die Treiber der rÃumlichen Emergenz und Konzentration der Photovoltaik-Industrie in Deutschland. *Zeitschrift fÃr Wirtschaftsgeographie*, 59(3).

Breyer, C., Birkner, C., Meiss, J., Goldschmidt, J. C., and Riede, M. (2013). A top-down analysis: Determining photovoltaics r&d investments from patent analysis and r&d headcount. *Energy Policy*, 62:1570 – 1580.

Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1):1–32.

Brunel, C. (2016). Green innovation and green manufacturing: Links between environmental policies, innovation and production. Technical report, Georgetown University.

Bruns, S. B. (2017). Meta-regression models and observational research. *Oxford Bulletin of Economics and Statistics*, 79(5):637–653.

Bruns, S. B. and Ioannidis, J. P. A. (2016). p-curve and p-hacking in observational research. *PLOS ONE*, 11(2):e0149144.

Bruns, S. B., König, J., and Stern, D. I. (2018). Replication and robustness analysis of 'energy and economic growth in the USA: A multivariate approach'. *Energy Economics*.

Bush, V. (1945). Science - the endless frontier. Technical report, A Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945.

Bäckström, K., Lundmark, R., and Söderholm, P. (2014). Public policies and solar PV innovation: An empirical study based on patent data. Technical report, IAEE International Conference 37.

Böhringer, C., Cuntz, A., Harhoff, D., and Otoo, E. A. (2014). The impacts of feed-in tariffs on innovation: Empirical evidence from Germany. Technical report, V-363-14, University of Oldenburg, Department of Economics (revised January 2014).

Calel, R. and Dechezleprêtre, A. (2016). Environmental policy and directed technological change: Evidence from the European carbon market. *Review of Economics and Statistics*, 98(1):173–191.

Cantner, U., Graf, H., Herrmann, J., and Kalthaus, M. (2016). Inventor networks in renewable energies: The influence of the policy mix in Germany. *Research Policy*, 45(6):1165–1184.

Chen, D.-Z., Dong, H.-R., and Huang, M.-H. (2010). A study of collaborations in solar cell science and technology. In *IEEE International Conference on Management of Innovation and Technology (ICMIT), 2010*, pages 794–799.

Chen, J. K. and Pham, V. K. (2014). A study on knowledge flows of dye-sensitized solar cells' patent. *Foresight*, 16(3):229–249.

Choe, H., Lee, D. H., Seo, I. W., and Kim, H. D. (2013). Patent citation network analysis for the domain of organic photovoltaic cells: Country, institution, and technology field. *Renewable and Sustainable Energy Reviews*, 26:492–505.

Choi, H. and Anadn, L. D. (2014). The role of the complementary sector and its relationship with network formation and government policies in emerging sectors: The case of solar photovoltaics between 2001 and 2009. *Technological Forecasting and Social Change*, 82:80–94.

Colombelli, A., Krafft, J., and Quatraro, F. (2015). Eco-innovation and firm growth: Do green gazelles run faster? Microeconomic evidence from a sample of european firms. Technical report, GREDEG WP No. 2015-12.

Corsatea, T. D. (2014). Technological capabilities for innovation activities across europe: Evidence from wind, solar and bioenergy technologies. *Renewable and Sustainable Energy Reviews*, 37:469 – 479.

Costantini, V., Crespi, F., and Curci, Y. (2015a). A keyword selection method for mapping technological knowledge in specific sectors through patent data: the case of biofuels sector. *Economics of Innovation and New Technology*, 24(4):282–308.

Costantini, V., Crespi, F., Martini, C., and Pennacchio, L. (2015b). Demand-pull and technology-push public support for eco-innovation: The case of the biofuels sector. *Research Policy*, 44(3):577–595.

Costantini, V., Crespi, F., Orsatti, G., and Palma, A. (2014). Policy inducement effects in energy efficiency technologies. an empirical analysis of the residential sector. In Ansuategi, A., Delgado, J., and Galarraga, I., editors, *Green Energy and Efficiency*, pages 201–232. Springer.

Costantini, V., Crespi, F., and Palma, A. (2015c). Characterizing the policy mix and its impact on eco-innovation in energy-efficient technologies. Technical report, SEEDS Working Paper Series 11/2015.

Costantini, V. and Mazzanti, M. (2012). On the green and innovative side of trade competitiveness? the impact of environmental policies and innovation on eu exports. *Research Policy*, 41(1):132–153.

Criscuolo, P. (2006). The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO. *Scientometrics*, 66(1):23–41.

Cumming, G. (2014). The new statistics - why and how. *Psychological Science*, 25(1):7–29. PMID: 24220629.

Dalmazzone, S. and Corsatea, T. D. (2012). A regional analysis of renewable energy patenting in Italy. *SSRN Electronic Journal*.

de la Tour, A., Glachant, M., and MÃ©niÃre, Y. (2011). Innovation and international technology transfer: The case of the chinese photovoltaic industry. *Energy Policy*, 39(2):761 – 770.

de Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., and van Pottelsberghe de la Potterie, B. (2013). The worldwide count of priority patents: A new indicator of inventive activity. *Research Policy*, 42(3):720–737.

de Rassenfosse, G., Jaffe, A. B., and Webster, E. (2016). Low-quality patents in the eye of the beholder: Evidence from multiple examiners. Working Paper 22244, National Bureau of Economic Research.

de Rassenfosse, G., Schoen, A., and Wastyn, A. (2014). Selection bias in innovation studies: A simple test. *Technological Forecasting and Social Change*, 81:287–299.

de Rassenfosse, G. and van Pottelsberghe de la Potterie, B. (2011). On the price elasticity of demand for patents. *Oxford Bulletin of Economics and Statistics*, 74(1):58–77.

de Rassenfosse, G. and van Pottelsberghe de la Potterie, B. (2013). The role of fees in patent systems: Theory and evidence. *Journal of Economic Surveys*, 27(4):696–716.

DechezleprÃªtre, A. and Glachant, M. (2014). Does foreign environmental policy influence domestic innovation? evidence from the wind industry. *Environmental and Resource Economics*, 58(3):391–413.

DechezleprÃªtre, A., Glachant, M., Haščič, I., Johnstone, N., and MÃ©niÃre, Y. (2009). Invention and transfer of climate change mitigation technologies on a global scale: A study drawing on patent data. Working Papers 2009.82, Fondazione Eni Enrico Mattei.

DechezleprÃªtre, A., Glachant, M., Haščič, I., Johnstone, N., and MÃ©niÃre, Y. (2011). Invention and transfer of climate change-mitigation technologies: A global analysis. *Review of Environmental Economics and Policy*, 5(1):109–130.

Dernis, H., Guellec, D., and van Pottelsberghe, B. (2001). Using patent counts for cross-country comparisons of technology output. Technical report, OECD.

Dernis, H. and Khan, M. (2004). Triadic patent families methodology. Technical report, OECD Science, Technology and Industry Working Papers 2004/02.

Diederich, H. and Althammer, W. (2016). Environmental regulation and innovation in renewable energy technologies. *SSRN Electronic Journal*.

Dirnberger, D. (2011). A guide to efficient keyword, sequence and classification search strategies for biopharmaceutical drug-centric patent landscape searches - a human recombinant insulin patent landscape case study. *World Patent Information*, 33(2):128–143.

Dong, H. R., Chen, D. Z., and Huang, M. H. (2013). The patterns of industry-university-government collaboration in photovoltaic technology. In *ISSI 2013: 14th International Society of Scientometrics and Informetrics Conference*, pages 1379–1391, Vienna, Austria.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, 67(4):559–579.

Eckert, A. and Langinier, C. (2014). A survey of the economics of patent systems and procedures. *Journal of Economic Surveys*, 28(5):996–1015.

Eisenschitz, T. and Crane, J. (1986). Patent searching using classifications and using keywords. *World Patent Information*, 8(1):38–40.

EPO (2014). Worldwide patent statistical database (patstat), april 2014 edition.

Frietsch, R. and Schmoch, U. (2010). Transnational patents and international markets. *Scientometrics*, 82(1):185–200.

FRINNOV (2009). Photovoltaic thin film cells. Technical report, France Innovation Scientifique & Transfert.

Gallagher, K. S. (2014). *The Globalization of Clean Energy Technology. Lessons from China.* The MIT Press.

Geng, J. B. and Ji, Q. (2016). Technological innovation and renewable energy development: evidence based on patent counts. *International Journal of Global Environmental Issues*, 15(3):217–234.

Glachant, M. and Dechezleprêtre, A. (2016). What role for climate negotiations on technology transfer? *Climate Policy*, pages 1–15.

Greene, W. H. (2012). *Econometric Analysis.* Pearson, New Jersey, 7th edition edition.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 28:1661–1707.

Groba, F. and Cao, J. (2015). Chinese renewable energy technology exports: The role of policy, innovation and markets. *Environmental and Resource Economics*, 60(2):243–283.

Grupp, H. (1996). Spillover effects and the science base of innovations reconsidered: an empirical approach. *Journal of Evolutionary Economics*, 6:175–197.

Grupp, H. and Schmoch, U. (1999). Patent statistics in the age of globalisation: new legal procedures, new analytical methods, new economic interpretation. *Research Policy*, 28(4):377–396.

Guan, J. C. and Yan, Y. (2016). Technological proximity and recombinative innovation in the alternative energy field. *Research Policy*, 45(7):1460–1473.

Guellec, D. and van Pottelsberghe de la Potterie, B. (2000). Applications, grants and the value of patent. *Economics Letters*, 69(1):109–114.

Haller, I. (2007). Mathematical modelling of innovation dynamics: An empirical analysis of the photovoltaic market in germany. Technical report, EPIP-2007 Conference material.

Han, X. and Niosi, J. (2016). Star scientists in PV technology and the limits of academic entrepreneurship. *Journal of Business Research*, 69(5):1707–1711.

Harhoff, D., Hoisl, K., Reichl, B., and van Pottelsberghe de la Potterie, B. (2009). Patent validation at the country level–the role of fees and translation costs. *Research Policy*, 38(9):1423–1437.

Harhoff, D., M.Scherer, F., and Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8):1343–1363.

Haščič, I. and Migotto, M. (2015). Measuring environmental innovation using patent data. Technical report, OECD Environment Working Papers No 89.

Hendry, D. F. (1980). Econometrics-alchemy or science? *Economica*, 47(188):387–406.

Herrmann, J. and Töpfer, S. (2017). Structural similarity and dependency of research networks in the German PV-industry. Mimeo, Friedrich Schiller University Jena.

Herrmann, J. K. (2017). Is there a ‘sailing ship’ effect for fossil fuels through renewable energies? Mimeo, Friedrich Schiller University Jena.

Hicks, J. R. (1932). *The Theory of Wages*. Macmillan, London.

Huang, M.-H., Chen, D.-Z., and Dong, H.-R. (2011). Identify technology main paths by adding missing citations using bibliographic coupling and co-citation methods in photovoltaics. In *Technology Management in the Energy Smart World (PICMET), 2011 Proceedings of PICMET '11:*, pages 1–6.

Huang, M.-H., Dong, H.-R., and Chen, D.-Z. (2013). The unbalanced performance and regional differences in scientific and technological collaboration in the field of solar cells. *Scientometrics*, 94(1):423–438.

Ioannidis, J. P. A., Stanley, T. D., and Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, 127(605):F236âF265.

Jaffe, A. B. and Trajtenberg, M. (2002). *Patents, Citations and Innovations: A window on the knowledge economy*. MIT Press, Cambridge, Massachusetts.

Jamali, M. Y., Aslani, A., Moghadam, B. F., Naaranoja, M., and Madvar, M. D. (2016). Analysis of photovoltaic technology development based on technology life cycle approach. *Journal of Renewable and Sustainable Energy*, 8(3):035905.

Jamasb, T. and Pollitt, M. G. (2011). Electricity sector liberalisation and innovation: An analysis of the UK's patenting activities. *Research Policy*, 40(2):309–324.

Jamasb, T. and Pollitt, M. G. (2015). Why and how to subsidise energy r+d: Lessons from the collapse and recovery of electricity innovation in the UK. *Energy Policy*, 83:197–205.

Jang, S.-L., Chen, L.-J., Chen, J. H., and Chiu, Y.-C. (2013). Innovation and production in the global solar photovoltaic industry. *Scientometrics*, 94(3):1021–1036.

Johnstone, N., Haščič, I., and Popp, D. (2017). Erratum to: Renewable energy policies and technological innovation: Evidence based on patent counts. *Environmental and Resource Economics*.

Johnstone, N., Haščič, I., and Popp, D. (2010). Renewable energy policies and technological innovation: Evidence based on patent counts. *Environmental and Resource Economics*, 45(1):133–155.

Kalthaus, M. (2016). Knowledge recombination along the technology life cycle. Jena Economic Research Papers 2016–012, Friedrich Schiller University Jena.

Kalthaus, M. (2017). Identifying technological sub-trajectories in photovoltaic patents. Jena Economic Research Papers 2017 - 010, Friedrich Schiller University Jena.

Kazuyuki, M. and Takanori, T. (2014). Differences in science based innovation by technology life cycles: The case of solar cell technology. Technical report, RIETI Discussion Paper Series 14-E-005.

Kruse, J. (2016). Innovation in green energy technologies and the economic performance of firms. Technical report, EWI Working Paper, No 16/02.

Kruse, J. and Wetzel, H. (2016). Energy prices, technological knowledge, and innovation in green energy technologies: A dynamic panel analysis of European patent data. *CESifo Economic Studies*, 62(3):397–425.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498.

Lanjouw, J. O. and Mody, A. (1996). Innovation and the international diffusion of environmentally responsive technology. *Research Policy*, 25(4):549–571.

Lanjouw, J. O., Pakes, A., and Putnam, J. (1998). How to count patents and value intellectual property: The uses of patent renewal and application data. *The Journal of Industrial Economics*, 46(4):405–432.

Leamer, E. E. (1985). Sensitivity analyses would help. *The American Economic Review*, 75(3):308–313.

Lee, K. and Lee, S. (2013). Patterns of technological innovation and evolution in the energy sector: A patent-based approach. *Energy Policy*, 59:415–432.

Lee, S., joo Lee, H., and Yoon, B. (2012). Modeling and analyzing technology innovation in the energy sector: Patent-based HMM approach. *Computers & Industrial Engineering*, 63(3):564–577. Energy Management and Economics.

Lei, X.-P., Zhao, Z.-Y., Zhang, X., Chen, D.-Z., Huang, M.-H., Liu, R.-S., Zheng, J., and Zhao, Y.-H. (2012). A study of technological collaboration in solar cell industry using patent analysis. In *Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET '12:*, pages 1034–1041.

Lei, X.-P., Zhao, Z.-Y., Zhang, X., Chen, D.-Z., Huang, M.-H., Zheng, J., Liu, R.-S., Zhang, J., and Zhao, Y.-H. (2013). Technological collaboration patterns in solar cell industry based on patent inventors and assignees analysis. *Scientometrics*, 96(2):427–441.

Lemley, M. A. and Sampat, B. (2012). Examiner characteristics and patent office outcomes. *Review of Economics and Statistics*, 94(3):817–827.

Leydesdorff, L., Alkemade, F., Heimeriks, G., and Hoekstra, R. (2015). Patents as instruments for exploring innovation dynamics: geographic and technological perspectives on photovoltaic cells. *Scientometrics*, 102(1):629–651.

Li, X., Zhou, Y., Xue, L., and Huang, L. (2014). Patent map for emerging technologies: A case of solar cells technology. In *Management of Engineering Technology (PICMET), 2014 Portland International Conference on*, pages 3558–3571.

Liu, J. S., Kuan, C.-H., Cha, S.-C., Chuang, W.-L., Gau, G. J., and Jeng, J.-Y. (2011). Photovoltaic technology development: A perspective from patent growth analysis. *Solar Energy Materials and Solar Cells*, 95(11):3130–3136.

Lizin, S., Leroy, J., Delvenne, C., Dijk, M., Schepper, E. D., and Passel, S. V. (2013). A patent landscape analysis for organic photovoltaic solar cells: Identifying the technology's development phase. *Renewable Energy*, 57:5–11.

Margolis, R. M. and Kammen, D. M. (1999a). Evidence of under-investment in energy r&d in the United States and the impact of federal policy. *Energy Policy*, 27(10):575 – 584.

Margolis, R. M. and Kammen, D. M. (1999b). Underinvestment: The energy technology and r&d policy challenge. *Science*, 285(5428):690–692.

Marin, G. and Lotti, F. (2017). Productivity effects of eco-innovations using data on eco-patents. *Ind Corp Change*, 26(1):125–148.

Marinova, D. and Balaguer, A. (2009). Transformation in the photovoltaics industry in Australia, Germany and Japan: Comparison of actors, knowledge, institutions and markets. *Renewable Energy*, 34(2):461–464.

Martinez, M. E. M., de Melo, A. A., and dos Reis, P. C. (2013). Technological mapping of the photovoltaic solar energy industry by patent documents deposited in Brazil. In *CONF-IRM 2013 Proceedings*.

McCloskey, D. N. and Ziliak, S. T. (1996). The standard error of regressions. *Journal of Economic Literature*, 34(1):97–114.

Montecchi, T., Russo, D., and Liu, Y. (2013). Searching in cooperative patent classification: Comparison between keyword and concept-based search. *Advanced Engineering Informatics*, 27(3):335–345.

Mowery, D. C. and Rosenberg, N. (1979). The influence of market demand upon innovation: A critical review of some recent empirical studies. *Research Policy*, 8(2):103–153.

Myong, S. Y. (2014). Recent patent issues on intermediate reflectors for high efficiency thin-film silicon photovoltaic devices. *Renewable and Sustainable Energy Reviews*, 37:90–99.

Nagaoka, S., Motohashi, K., and Goto, A. (2010). Patent statistics as an innovation indicator. In Hall, B. H. and Rosenberg, N., editors, *Handbook of the Economics of Innovation*, volume 2 of *Handbook of the Economics of Innovation*, chapter 25, pages 1083–1127. North-Holland.

Nesta, L., Vona, F., and Nicolli, F. (2014). Environmental policies, competition and innovation in renewable energy. *Journal of Environmental Economics and Management*, 67(3):396–411.

Noailly, J. (2012). Improving the energy efficiency of buildings: The impact of environmental policy on technological innovation. *Energy Economics*, 34(3):795–806.

Noailly, J., Batrakova, S., and Lukach, R. (2010). Home, green home - A case study of inducing energy-efficient innovations in the Dutch building sector. CPB Documents 198, CPB Netherlands Bureau for Economic Policy Analysis.

Noailly, J. and Ryfisch, D. (2015). Multinational firms and the internationalization of green r&d: A review of the evidence and policy implications. *Energy Policy*, 83:218–228.

Noailly, J. and Shestalova, V. (2017). Knowledge spillovers from renewable energy technologies: Lessons from patent citations. *Environmental Innovation and Societal Transitions*, 22:1–14.

Noailly, J. and Smeets, R. (2012). Directing technical change from fossil-fuel to renewable energy innovation: An empirical application using firm-level patent data. Technical report, CPB Discussion Paper — 237.

OECD (2009). *OCED Patent Statistics Manual*. OECD, Paris.

Ordover, J. A. (1991). A patent system for both diffusion and exclusion. *Journal of Economic Perspectives*, 5(1):43–60.

Palangkaraya, A., Webster, E., and Jensen, P. H. (2011). Misclassification between patent offices: Evidence from a matched sample of patent applications. *Review of Economics and Statistics*, 93(3):1063–1075.

Patel, C. J., Burford, B., and Ioannidis, J. P. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, 68(9):1046–1058.

Pavitt, K. (1985). Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics*, 7(1-2):77–99.

Peters, M., Schmidt, T. S., Wiederkehr, D., and Schneider, M. (2011). Shedding light on solar technologies – a techno-economic assessment and its policy implications. *Energy Policy*, 39(10):6422–6439.

Peters, M., Schneider, M., Griesshaber, T., and Hoffmann, V. H. (2012). The impact of technology-push and demand-pull policies on technical change â Does the locus of policies matter? *Research Policy*, 41(8):1296–1308.

Pillu, H. (2009). *Knowledge flows through patent citation data*. PhD thesis, ÃCOLE CENTRALE PARIS.

Pillu, H. and KolÃ©da, G. (2009). Induced innovation and international technological opportunity in the field of energy: Evidence from world patent citations. Technical report, ERASME Ecole Centrale Paris.

Popp, D. (1997). *Induced Innovation, Energy Prices, and the Environment*. PhD thesis, Yale University.

Popp, D. (2016). From science to technology: The value of knowledge from different energy research institutions. Working Paper 22573, National Bureau of Economic Research.

Popp, D., Haščič, I., and Medhi, N. (2011). Technology and the diffusion of renewable energy. *Energy Economics*, 33(4):648–662.

Popp, D., Santen, N., Fisher-Vanden, K., and Webster, M. (2013). Technology variation vs. r&d uncertainty: What matters most for energy patent success? *Resource and Energy Economics*, 35(4):505–533.

Popp, D. C. (2001). The effect of new technology on energy consumption. *Resource and Energy Economics*, 23(3):215–239.

Porter, A., Youtie, J., Shapira, P., and Schoeneck, D. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research*, 10(5):715–728.

Putnam, J. (1996). *The Value of International Patent Rights*. PhD thesis, Yale University.

Rai, V., Metteauer, M., Querejazu, D., Wise, R., and Hamilton, G. (2013). Demand-pull and innovation in the U.S. solar market. Technical report, Working Paper Draft: July 29, 2013.

Sakakibara, M. and Branstetter, L. (2001). Do stronger patents induce more innovation? Evidence from the 1988 Japanese patent law reforms. *RAND Journal of Economics*, 32(1):77–100.

Schankerman, M. and Pakes, A. (1986). Estimates of the value of patent rights in european countries during the post-1950 period. *The Economic Journal*, 96(384):1052–1076.

Schmookler, J. (1966). *Invention and economic growth*. Harvard University Press, Cambridge.

Shibata, N., Kajikawa, Y., and Sakata, I. (2010). Extracting the commercialization gap between science and technology – case study of a solar cell. *Technological Forecasting and Social Change*, 77(7):1147–1155.

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, 22(11):1359–1366.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2):534–547.

Squicciarini, M., Dernis, H., and Criscuolo, C. (2013). Measuring patent quality - indicators of technological and economic value,. Technical report, OECD Science, Technology and Industry Working Papers 2013/03.

Stanley, T. D. and Jarrell, S. B. (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys*, 3(2):161–170.

Stek, P. E. and van Geenhuizen, M. (2015). Mapping innovation in the global photovoltaic industry: A bibliometric approach to cluster identification and analysis. Technical report, ERSA 55th Congress, World Renaissance: Changing roles for people and places, Lisbon, Portugal, 25-28 August 2015.

Sternitzke, C. (2009). Defining triadic patent families as a measure of technological strength. *Scientometrics*, 81(1):91–109.

Tseng, C.-Y. (2014). A comparative study of country level in technological innovation of six renewable energies. *International Journal of Economics and Research*, 5(6):27–37.

Tseng, F.-M., Hsieh, C.-H., Peng, Y.-N., and Chu, Y.-W. (2011). Using patent data to analyze trends and the technological strategies of the amorphous silicon thin-film solar cell industry. *Technological Forecasting and Social Change*, 78(2):332–345.

van Pottelsberghe de la Potterie, B. (2011). The quality factor in patent systems. *Industrial and Corporate Change*, 20(6):1755–1793.

Venugopalan, S. and Rai, V. (2015). Topic based classification and pattern identification in patents. *Technological Forecasting and Social Change*, 94:236–250.

Vijvers, W. (1990). The international patent classification as a search tool. *World Patent Information*, 12(1):26–30.

Wang, M.-Y. and Chiu, T.-F. (2010). Monitoring the key technology trends by combining chance discovery and survival analysis: Study on solar cell patent documents. In *Technology Management for Global Economic Growth (PICMET), 2010 Proceedings of PICMET '10:*, pages 1–7.

Wangler, L. U. (2013). Renewables and innovation: Did policy induced structural change in the energy sector effect innovation in green technologies? *Journal of Environmental Planning and Management*, 56(2):211–237.

Watanabe, C., Tsuji, Y. S., and Griffy-Brown, C. (2001). Patent statistics: deciphering a 'real' versus a 'pseudo' proxy of innovation. *Technovation*, 21(12):783–790.

Wilholt, T. (2009). Bias and values in scientific research. *Studies in History and Philosophy of Science Part A*, 40(1):92–101.

Wilholt, T. (2013). Epistemic trust in science. *The British Journal for the Philosophy of Science*, 64(2):233–253.

Wong, C.-Y., Keng, Z.-X., Mohamad, Z. F., and Azizan, S. A. (2016). Patterns of technological accumulation: The comparative advantage and relative impact of Asian emerging economies in low carbon energy technological systems. *Renewable and Sustainable Energy Reviews*, 57:977–987.

Wong, C.-Y., Mohamad, Z. F., Keng, Z.-X., and Azizan, S. A. (2014). Examining the patterns of innovation in low carbon energy science and technology: Publications and patents of Asian emerging economies. *Energy Policy*, 73:789–802.

Wu, C.-Y. (2014). Comparisons of technological innovation capabilities in the solar photovoltaic industries of Taiwan, China, and Korea. *Scientometrics*, 98(1):429–446.

Wu, C.-Y. and Hu, M.-C. (2015). The development trajectory and technological innovation capabilities in the global renewable energy industry. In *2015 Portland International Conference on Management of Engineering and Technology (PICMET)*. Institute of Electrical & Electronics Engineers (IEEE).

Wu, C.-Y. and Mathews, J. A. (2012). Knowledge flows in the solar photovoltaic industry: Insights from patenting by Taiwan, Korea, and China. *Research Policy*, 41(3):524 – 540.

Wurlod, J.-D. and Noailly, J. (2016). The impact of green innovation on energy intensity: An empirical analysis for 14 industrial sectors in oecd countries. Technical report, Research Paper 42 — 2016.

Xie, Z. and Miyazaki, K. (2013). Evaluating the effectiveness of keyword search strategy for patent identification. *World Patent Information*, 35(1):20–30.

Zachmann, G., Serwaah-Panin, A., and Peruzzi, M. (2015). When and how to support renewables?–Letting the data speak. In Ansuategi, A., Delgado, J., and Galarraga, I., editors, *Green Energy and Efficiency*, Green Energy and Technology, pages 291–332. Springer International Publishing.

Zheng, C. and Kammen, D. M. (2014). An innovation-focused roadmap for a sustainable global photovoltaic industry. *Energy Policy*, 67:159–169.