Patent Quality, Intellectual Property Rights, and Technology Transfer in the Solar Sector: All in the Family?

by

Dr. Amanda Phalin, Lecturer
Department of Finance, Insurance and Real Estate
University of Florida, Warrington College of Business
BRY 233A, PO Box 117168
Gainesville, FL 32611-7168
Tel: 352.392.6527
Fax: 352.392.0301

Abstract

This paper investigates the relationship between patent quality and the international transfer of solar technology using patent filings from the U.S. to 22 high-income countries worldwide. The data are examined with two different methods: analyzing patent equivalents and analyzing extended patent families. The results show that when examining patent equivalents: (1) a positive and statistically significant relationship exists between patent quality and the international transfer of solar technology and (2) contrary to other research, intellectual property rights (IPR) laws alone generally have no effect or a negative effect on technology transfer in this sector when a quality measure is included. However, when examining extended patent families: (1) generally no relationship is found between patent quality and technology transfer in the solar industry, but (2) a positive, statistically significant relationship between IPR laws and patents filed occurs. These varying results have significant implications for governments attempting to craft policies that increase the transfer and adoption of green technologies, and should encourage researchers to consider carefully how patent flows are defined and used to measure technology transfer.

Keywords: Intellectual property rights, patents, patent citations, patent equivalents, patent families, solar technology, green technology, technology transfer.

JEL Classification: F64, O33, O34, Q42, Q55.

Acknowledgments: I would like to thank Dr. Elias Dinopoulos, Dr. Walter Park, Dr. Adam Narkiewicz, and Joe Whitman for their helpful comments and suggestions. I also thank Dr. Adam Narkiewicz and Joe Whitman for their assistance with data collection and management.

1: Introduction

Much discussion has recently focused on international environmental technology transfer and its role in facilitating climate-change-mitigating technology, especially in emerging and lesser-developed markets. Of particular concern is the role of patents in both measuring and facilitating these flows. Key questions emerge: Which patents are likely to be filed abroad, and what characteristics do they share? What do countries with high patent inflows—and high-quality patent inflows—look like? How can researchers best investigate these phenomena?

Using patent flows as a measure of technology diffusion, as well as a weighted measure of forward patent citations to gauge quality, I investigate whether higher-quality solar patents are more likely to be filed in countries with more stringent intellectual property rights (IPR) laws, more developed infrastructure, and/or higher levels of human capital. I do this using two different methods: First, I examine patent equivalents from the U.S. filed abroad, and then I examine extended patent families originating in the U.S. and filed abroad.

This issue is especially salient now, given the 2015 U.N. climate conference in Paris, known as COP 21, where both developed and developing countries committed themselves to moving toward an emissions-free future (WRI, 2015). At the 2011 Durban conference, attendees formulated a legally-binding agreement to reduce climate-change-causing emissions, and Kyoto Protocol policies were extended (WRI, 2011). Both conferences addressed financing for climate-change-mitigating technologies in poorer nations, and facilitating technology transfer is important in this context. Thus, solar technology transfer is a significant area of research. Maskus (2016) notes that between 1997 and 2008 alone, global patenting in environmental technologies grew about 20 percent per year. In addition, solar investment rose almost 30 percent between 2013 and 2014 (Meléndez-Ortiz, 2016), while solar photovoltaic (PV) capacity increased 25 percent between 2014 and 2015 alone (REN21, 2016). Overall, the size of the solar PV market is 10 times larger than it was a decade ago (REN21).[1] In other words, these technologies are becoming more and more widespread.

This research adds to the literature in several key ways. First, I examine data disaggregated at the technology level. While much research has explored the relationship between intellectual property and technology transfer at the aggregate level, "there is an urgent need for increased availability of reliable and objective data on climate technologies, particularly on IPR-related aspects" (Latif and Maskus, et al., 2011).[2] Second, I add to the literature by comparing two different methods of analyzing patent flows:

[1] The top three markets for solar PV are China, Japan, and the U.S. The top two markets for solar concentrating power are Spain and the U.S.

[2] Other scholars have previously pointed out the need to examine data in this area that is disaggregated by national income level and sector, including Kumar (1996), Lall (2003), and Basberg (1987).

equivalents vs. extended families. Definitions of these categories can vary from organization to organization (Adams, 2012). Martínez (2006) found that differences in definitions may lead to different results depending on which data are analyzed; ergo, considering patent equivalents vs. patent families at a disaggregated level is vital. To my knowledge, this is the first study of its kind conducted for the solar sector, or indeed any environmental technology.

Quality is a fundamental aspect of the IPR–technology-transfer question. Theory predicts that higher-quality technologies are more likely to be diffused internationally (Eaton and Kortum, 1994; Eaton and Kortum, 1995, Kortum and Lerner, 1997). How is quality best measured? Beginning in the 1980s, scholars used patent citations for this purpose. Citations come in two types: backward and forward. Every time a patent is filed, the author and patent office officials list prior art, which includes previous patents that may be similar to or relevant to the current patent-filer's technology. Let the current patent being filed = X. We can refer to all the prior art contained in X as backward patent citations of X. However, going forward, if other patents cite X as prior art, we can refer to them as forward citations of X.

Researchers like citations because they provide a clear path showing how innovation "moves" between people, firms, industries, and countries. Citations can be indicative of a product's quality because the more often a patent is cited, the higher the probability that it is a useful or valuable development. I use a weighted measure of forward patent citations to gauge patent quality. I hypothesize that, per theory, patent quality should be positively correlated with more technology diffusion.

IPRs are also important for green technologies since this sector requires large initial R&D investments; innovators need to be able to reap profits from their initial outlays to succeed (Latif and Maskus, et al., 2011). In addition, IPRs' importance can vary across technologies; some sectors that produce easily imitated products (e.g., pharmaceuticals) find IPR strength to be a vital requirement, while other, less-imitable, industries (e.g., traditional manufacturing) may be less concerned with protecting intellectual property.

I expect that, because of the large R&D investment required and the advanced technology often needed to develop and produce solar-energy products, a strong IPR system should be positively correlated with patent inflows. The only study, to my knowledge, examining the relationship between IPR laws and solar international technology transfer (ITT) (Dechezleprêtre, 2011) found that stronger IPR laws have a positive, statistically significant effect on solar technology flows.

By examining patent and patent quality data in the solar sector, I hope to be able to explain what kinds of systems—economic, legal, and otherwise—lesser developed countries may need to have in place in order to attract high-quality solar technology investment. The paper proceeds as follows: Section 2 presents a brief review of the related literature. Section 3 discusses differences between patent equivalents and extended families and how they may affect results. Section 4 explains the theoretical underpinnings of the analysis, as well as an econometric model drawn from the theoretical model. Section 5 details how the variables are constructed and the sources of data. Section 6 presents and analyzes results for the patent equivalent analysis. Section 7 presents and analyzes results for the extended patent families analysis. Section 8 offers concluding remarks.

2: Related Literature

Little has been written about patent quality as it specifically relates to the international transfer of environmental technology. However, the literature on patent quality, patent citations, intellectual property, and technology diffusion in general is well developed. These studies attempt to answer some key questions. First, how related are citations to quality? Second, does quality predict patent filings abroad? The literature dealing specifically with environmental technologies and IPRs examines what effect strong patent laws may have on technology inflows.

Dozens of studies have used forward patent citations to measure patent quality. A 2001 NBER analysis of worldwide patent data (of which the dataset in this study is a subset) showed that forward patent citations occur over long periods of time. Specifically, 50 percent of patents will receive citations within 10 years of filing, 25 percent more will receive citations within 20 years of filing, and 5 percent more will receive citations within 50 years or more after filing. This means that if newer patents are included in the data, they most likely will not reflect the correct forward-citation effect simply because these patents are not old enough to have received all of the citations they will likely garner (Hall et al., 2001). Van Zeebroeck (2011) offers a simpler remedy to this time issue: Count citations received by patent applications within a given period of time.

Citations received may also differ by technology. For instance, Dechezleprêtre, Martin and Mohnen (2013) find that "clean," or environmentally friendly technologies, in the energy production, cars, fuel, and lighting sectors receive on average 43 percent more citations than "dirty," or non-green, patents. This difference can be accounted for by two factors: 1) clean inventions have wider, more general uses, and 2) they are more novel compared with older "dirty" technologies. Therefore, focusing on a specific sector can help ameliorate some of these problems. Indeed, Popp (2006) finds empirical support for the idea that "allowing for different behavior across technologies is important" for climate-change-mitigating technologies. Thus, one can conclude that examining data sector by sector, or technology by technology, is apt to yield the most accurate results.

Although a fairly rich literature covers patent citations and patent quality[3], very little empirical work has examined these issues for green technologies. Haščič, I. and M. Migotto (2015) analyze different methods for measuring innovation in environmental technologies and conclude that "patent data are best suited for identifying specifically 'environmental' innovations." Acosta et al. (2009) examines European environmental patents and using weighted citations as a quality measure. The authors find that patents from institutions are of a higher quality than those from individuals. They also find that green patents from the United States and Japan are cited more frequently than European patents. Finally, their analysis shows that patents that can be used in multiple sectors are more likely to be cited than patents that have very specific, limited uses. Jaffe and de Rassenfosse (2016) also note that the most highly cited patents generally are more likely to be from the U.S., and more specifically from U.S. corporations.

Another recent study of patent citations and environmental technology (Pillu and Koléda 2009) examines 11 energy technologies in France, Germany, Japan, the United Kingdom, and the United States to determine what factors induce innovation in this industry. The authors use patent citations to help construct a proxy for the available stock of knowledge that inventors can use to develop new innovations; they weight the stock of patents by their productivity, i.e., citations. The authors find that both high energy prices and the availability of knowledge (i.e., patent citations) encourage innovation.

The literature confirms a robust positive relationship between patent citations and patent quality. This allows us to explore another key issue: Is there also a positive correlation between patent quality and patent filings? The available research shows the answer to be "maybe." Little research has been done on patent quality and patent flows; however, several studies have explored the relationship between patent quality and patent valuation. One of the earlier studies in this area (Scherer 1984) showed that for U.S. firms, higher-quality patents are worth more. Later studies have also confirmed that high-quality patents are also worth more, including Hirschey and Richardson (2001, 2004) and Lanjouw and Schankerman (2004). Chen and Chang (2010) find that in the U.S. pharmaceutical industry, only some indicators of quality are positively associated with firm value. Previous work has shown a correlation between forward citations and patent value (Jaffe & de Rassenfosse, 2016). Lanjouw and Schankerman (1999) find that among U.S. manufacturing firms, higher-quality patents are more likely to be renewed, and firms are more likely to sue when high-quality patents are infringed upon. Jaffe and de Rassenfosse note that "the relevance of patent data as innovation indicator … varies across fields." Thus, it is again important to examine disaggregated data at the technology level, as done in this research.

[3]See Trajtenberg (1990), (Lanjouw, 2004), Harhoff (2003), Marco (2007), Harhoff et al. (1999), Fallah et al. (2009), Rosenkopf and Nerkar (2011), Norback et al. (2011), Hall et al. (2005), and Acosta et al. (2009), all listed in the "References" at the end of this paper.

The second question this research attempts to answer is whether strong IPR laws in general foster higher technology flows. The results of previous studies analyzing this issue are mixed. Using R&D expenditures as the dependent variable, Kunwar and Evenson (2009) found “at best, weak evidence” supporting the claim that stronger IPR laws lead to an increase in technology. They do note, however, that countries with weak laws may also experience low levels of human capital and economic growth, which can also hinder technological development. Phalin (2011) noted that increased IPR restrictions can attract more technology in middle-income and high-income nations, but may have no or negative impact for very poor countries. This is why it is important to examine these issues not only disaggregated by industry, but also by income level of countries. Adams (2010) found that although increasing IPR protections does have a positive effect on foreign direct investment in developing countries, openness and growth in the economy and investment are also important drivers.

It is also useful to discuss the literature specifically related to intellectual property rights protection and green technologies. Namely, do stronger IPR protections engender more environmental innovation? Barton (2007) makes one of the first attempts to examine the relationship between IPR protection and environmentally sound technologies (ESTs). In photovoltaics, he concludes that patents might not present an obstacle to access for developing nations due to the level of competition induced by the high number of businesses in the industry worldwide. Similarly, Hall and Helmers (2010) note that patenting in climate-change-mitigating technologies is less concentrated than patenting in general, with China (rather than just Germany, Japan, and the U.S.) also producing a small percentage of total green patents. Dechezleprêtre, Glachant, Ménière (2010) analyze data from 66 countries between 1990 and 2003 to determine whether higher IPR protection increases the transfer of ESTs. The authors find a statistically significant, positive relationship between the strength of a country’s IPR laws and patents filed in wind, solar, hydro, cement, building, and methane. They find no statistically significant relationship in biomass, geothermal, waste, and fuel injection. They find a statistically significant negative relationship in ocean and light.

Finally, researchers must also consider country and industry characteristics, which can affect technology diffusion as well. Almost all studies exploring the relationship between intellectual property and technology diffusion use a set of independent variables to control for national factors that may affect the decision to patent (Branstetter et al., 2006, 2007; Evenson and Kanwar, 2001; Javorcik, 2004; Kanwar, 2009; and Maskus et al., 1995, 2001, 2005, 2005). To control for market size, researchers may use GDP, per capita GDP, or population. When dealing with innovation diffusion, it is also essential to measure a country’s capacity to absorb new technologies; various measures of human capital are used, including years of secondary or tertiary education, or the population employed in high-tech or R&D sectors. Studies may also account for a nation’s economic relationship to the rest of the world, so they might control for membership in a trade bloc or other trade agreements. These studies have also found that controlling for industry can yield better results. For example, researchers have examined the different effects IPRs can have in traditional manufacturing vs. more high-tech sectors such as chemicals and pharmaceuticals (Javorcik, 2004). Overall, it is important to consider a wide array of factors in addition to patent quality and IPR laws that may affect patent flows.

3: Patent Equivalents vs. Patent Families

No legal definitions of patent families vs. patent equivalents exists; rather, each database provider (e.g., USPTO, EPO, WIPO, OECD) creates its own definition, which can vary from organization to organization (Adams, 2012). Martínez (2011) found that differences in definitions may lead to different results for 25 percent of patent families with "complex" structures (i.e., patents with multiple first filings and/or those with indirect subsequent filings), and that these differences may affect sectors differently.

The European Patent Office defines patent equivalents as documents whose priorities are the same; in other words, the same declaration of priority, or patent document, has been filed in more than one place. This is also known as patent-to-patent or a "simple patent family" (Intellogist, 2016; EPO, 2016). The results discussed in Section 6 are derived using this method. Martínez (2011) notes that "equivalents are widely seen as the preferred definition when trying to identify the legal duplicates of an application in other offices." However, she also writes that it remains unclear whether all those individual patents are actually related to the same overall invention, and that more research is needed in this area.

The results described in Section 7 are derived from examining extended patent families. This means "a patent family relating to one or more inventions, each member of which has for the basis of its 'priority right' at least one originating application in common with at least one other member of the family" (Intellogist, 2016). The European Patent Office calls this definition "much broader" because it "includes documents having the same scope but lacking a common priority" (EPO, 2016). Indeed, according to the EPO, "In the European Patent Register a patent family is defined as comprising all the documents sharing – directly or indirectly (e.g. via a third document) – at least one priority. This includes all the patent documents resulting from a patent application submitted to a patent office as a first filing and from the same patent application filed within the priority year with a patent office in any other country" (Espacenet, 2016). This wider definition means that an extended patent family captures patents that are not only directly related to, but also indirectly related to, the original priority document.

Figure 1 shows the difference between equivalents/simple families vs. extended families. D2–D3 are equivalents, or a simple family, because they share the same priorities, P1 and P2. The extended families, however, are much broader and include documents that are only indirectly linked to one another. Note, for example, that D2-D4 are in the same family (P2) because they share at least one priority with another member, but that only D2 and D3 are associated with P1 but D4 is not. Yet, because of this indirect relationship, D2 and D4 are in the same extended family. Indeed, research has shown that extended families may cover different inventions that are related only by sector (Mailänder, 2012).

Keeping these definitions in mind, arguments can be made for using both methods. Equivalents compare patent-to-patent, so researchers can follow the trail of a small piece of technology around the world, from patent office to patent office. However, products as a whole rarely contain one patent; they can have a few or dozens of individually patented components that comprise the whole. If the intent is to track the transfer of a product, complete invention, or complete technology, rather than a component technology, then the extended patent family definition makes more sense. For that same reason, using equivalents may result in the over-counting of complete technologies, because several patents may be used as component parts of a finished invention (Martínez, 2011). By using extended families, however, "it is possible to find two patent documents with no priority in common, but which are indirectly related because they both share at least one priority with a third application" (Martínez, 2011). In other words, those two patents may be defined as being related but in reality are really not; therefore, the extended family not accurately track one technology or invention. Both methods raise issues; therefore, in this paper, I analyze solar technology flows using both patent equivalents and extended families.[4]

4: Theoretical Framework

This work is based on the model of Gallini et al. (2006), who base their work on Eaton and Kortum (1994, 1995), Kortum and Lerner (1997), and Rafiquzzaman and Whewell (1998). Gallini et al. analyze aggregate data of patents filed in Canada from Germany, the United Kingdom, and the United States. Eaton and Kortum (1994) model the creation of new inventions and their international diffusion. In their model, the value of a patent depends on its quality, q, a random variable drawn from a cumulative distribution. They derive the following threshold condition:

$$V_{nit}^{pat}(q) - V_{nit}^{not}(q) = c_{nit} \quad (1)$$

$V_{nit}^{pat}(q)$ is the value of filing a patent with quality q from country i in country n; $V_{nit}^{not}(q)$ is the value of not filing a patent with quality q from country i in country n. The patent will be filed as long as $c_{nit} > 0$. Three country characteristics directly affect this threshold: the lag time it takes for the technology to be adopted in country n, the strength of patent protection laws in country n, and the cost of patenting in country n. These can be proxied empirically by a measure of human capital, an index of patent rights, and filing fees or the need for translation, respectively.

[4] The two datasets used to analyze the two different methods contain other small differences in independent variables, which will be discussed in Section 5.

In a later version of this paper (Eaton and Kortum, 1995), the authors expand the model to include the determinants of technology diffusion, "i.e., the probability that an invention from country i will be adopted in country n. We let diffusion from country i to country n depend on: (1) whether n and i are the same country or not, (2) the distance between n and i, (3) the level of human capital in n (the adopting country), and (4) the level of country n's imports from i relative to n's GNP."

Gallini et al. (2006) follow Eaton and Kortum to derive a model measuring the propensity to patent. Their specification is as follows:

$$E\left(\frac{P_{ij}}{N_i^*}\right) = \rho(s_j, x_j, z_{ij}, c_j) \qquad (2)$$

Where P_{ij} is the number of patents filed in destination country j by the source country i; N_i^* is the innovation effect, or the total number of patentable inventions (which is unobservable); ρ is the probability that an invention from country i will be high quality enough for the patent filing to be profitable in country j; s_j is the strength of patent protection in country j; x_j is a set of indicators controlling for the economic environment in j (i.e., GDP, human capital); z_{ij} is a set of indicators describing the relationship between i and j (i.e., distance, trade flows); and c_j is the cost of filing a patent in country j. Taking logs, their econometric specification is as follows:

$$log\left(\frac{P_{ijt}}{n_{it}}\right) = \alpha_0 + \beta s_{jt} + \gamma x_{jt} + \delta z_{ijt} + \theta c_{jt} + \alpha_t + \alpha_i + \alpha_{it} + \epsilon_{ijt} \qquad (3)$$

As above, P_{ijt} is the number of patents from the source country, i, filed in the destination country, j (Canada), in year t. n_{it} is the amount spent on R&D in i in year t. s_{jt} is the strength of patent protection in j as measured by the Ginarte and Park Index (Ginarte and Park, 1997). x_{jt} is a set of variables measuring human capital, GDP, and an index measuring the effectiveness of j's antitrust laws. z_{ijt} describes the relationship between i and j, including: distance, distance squared, and log(j's imports from i / real GDP). c_{jt} controls for the cost of patenting in j, including fees and a dummy variable indicating whether translation is required. Finally, the authors include time and country fixed effects.

I modify the Gallini et al. model in several ways. First, in this model—as in those of Eaton and Kortum and Kortum and Lerner—quality is randomly drawn from a distribution. I add a quality variable on the right-hand side: a weighted measure of the total number of patent citations from i in year t. Second, I perform a disaggregated analysis, breaking down the patent data and examining only solar technology. This is important because not all results will be the same across industries and technologies. The Gallini specification would then be as follows:

$$E(P_{ijt}) = \rho(s_j, x_j, z_{ij}, c_j, q_{ij}) \qquad (4)$$

$$log(P_{ijt}) = \alpha_0 + \beta \log(s_{jt}) + \gamma \log(x_{jt}) + \delta \log(z_{ijt}) + \theta \log(c_{jt}) + \varphi \log(q_{ijt}) + \epsilon_{ijt} \qquad (5)$$

In my analysis, the source country, i, is the United States. I will use 22 nations as destination countries. In addition, I add a variable controlling for the existence of pro-renewable-energy policies in country j.

Gallini et al. use a log-linear specification to measure the propensity to patent. However, the dependent variable in this case is a count, and a significant portion of them are zeros. It can be argued that OLS specifications are better matched to continuous, as opposed to discrete, data. Moreover, when data on the dependent variable contain a large portion of zeros, as do the data here, it may be better to use a model that takes this into account. The first model to consider using count data is the Poisson model. However, Poisson requires $E(y|x) = Var(y|x)$, i.e., that the mean equals the variance. This is unlikely in the current case; thus, a negative binomial specification, which allows and corrects for differences in the variance, should be better suited to this analysis. Indeed, the large χ^2 values in all the results indicate that the data are not Poisson, that they are overdispersed, and that a negative binomial specification is appropriate.

5: Variables, Data, and Econometric Specification

The dataset includes the United States as the source country of solar patents and 22 nations as destination countries where the patents may be filed. The dependent variable, P_{ijt}, is the number of solar patents filed in country j from country i (the United States) in year t. Data for this variable were downloaded from Espacenet, the patent database of the European Patent Office.[5] Note that due to the model's specification, this is an aggregate measure of the number of solar patents, rather than an examination of individual patents. To gather the relevant information required for this study, I assembled all the International Patent Classification (IPC) codes that relate to solar technologies from the World Intellectual Property Organization's (WIPO) "IPC Green Inventory" list, which was created to allow researchers to identify environmentally sound technologies more easily.

Although the data used to examine patent equivalents and extended patent families are very similar, there are some small differences. First, the dependent variable in the equivalents dataset is from 1954–2011, downloaded in 2011. The dependent variable for the extended families dataset is expanded to include 1954–2015, downloaded in 2015. In addition, the list of solar-related IPC codes expanded from 92 in 2011 to 148 in 2015. This is because new technologies are being created at a rapid pace, and some require completely new coding; a good example of this is the recent development of solar nanotechnology, which is on the bleeding edge of R&D in this sector. Other minor differences in the independent variables are noted below.

Variables of interest:

[5] Information about the data-collection process for both methods used in this paper can be found in Appendices 1 and 2.

qual: A weighted proxy for the aggregate quality of the patents being filed in country j from the United States in year t equal to the total number of citations received by all patent equivalents or patent families from the U.S. filed in country j in year t divided by the total number of patents filed from the U.S. in country j in year t. As theory predicts, I expect the coefficient on *qual* to be positive.

ipr: The Ginarte and Park index, which measures "how strongly patent rights will be protected" in a given country (Ginarte and Park, 1997).[6] The scale is measured from 0 to 5, with 0 representing the weakest patent laws and 5 representing the strongest patent laws. Because of large R&D investments usually required for soalr technology, I expect the coefficient on *ipr* to be positive.

Control Variables:

gdp: Taken from the Penn World Tables, this variable controls for market size. For the patent equivalent analysis, it is total PPP converted GDP in millions of 2011 dollars. For the extended patent families analysis, it is real expenditure base GDP per capita in 2005 dollars (Heston et al., 2011). I hypothesize that larger markets will be more likely to draw patents because there is a higher chance of profitability in more developed economies. Thus, I expect the coefficient on *gdp* to be positive.

humk: A measure of human capital that predicts the destination country's ability to absorb new technologies and innovations. A larger stock of human capital will signal that a country is better equipped to deal with new technologies. I expect a higher level of patenting in countries with more human capital, and thus a positive coefficient on the *humk* variable. These data are found in the Barro and Lee dataset on worldwide educational attainment (Barro and Lee, 2013).

dist: A measure of direct-line distance between Washington, D.C., and country j's capital, per Gallini et al. These distances can be found in Fitzpatrick and Modlin (1986). Because countries that are nearer to one another tend to have higher trade and closer economic relationships, I expect the coefficient on this variable to be negative; i.e., the farther the distance, the fewer patents filed.

imps: Bilateral trade flows; countries with higher trade flows exchange more products and technology, so I expect the coefficient on this variable to be positive. For the patent equivalent analysis, the data for *imps* comes from the Feenstra and Lipsey NBER–United Nations Trade Data 1962–2000 dataset. For the extended patent families analysis, the data for *imps* comes from United Nations Comtrade Database.

lang: This variable controls for the cost of filing a patent in country j. For the equivalents analysis, a dummy variable equals 1 if translation is required, i.e., if the source country (U.S.) and destination country *do not share* a common language. For the extended families analysis, a dummy variable equals 1 if the source and destination countries *do share* an official language (i.e., if translation of the patent is not required). Controlling for the cost of filing a patent is difficult because so little data exist. I could attempt

[6] I thank Dr. Walter Park, who generously provided me the latest edition of the index, which includes rankings up to 2010.

to find information on current patent costs in each of the countries in the dataset and assume that costs are constant over time, as did Gallini et al. However, since the 1980s, patent costs have risen in Japan and the United States, while they have fluctuated at the European Patent Office (de Rassenfosse and van Pottelsberghe, 2012). Data for many nations are not readily available.

Another option is to exploit the fact that the cost of translation fees for patents can range in the thousands of dollars and therefore represent a significant portion of the overall cost of filing patents abroad (European Commission, 2010). There is no way to obtain specific information on translation fees since they are generally done by private companies; however, the dummy variable, while a blunt measure, does have the advantage of bringing more national specificity to the analysis. Although not an ideal gauge of cost, excluding a cost measure would be worse for the analysis than having no measure at all. I expect the coefficient on the variable *lang* to indicate that sharing a language leads to increased patent flows. Information for this variable was found in the CIA World Factbook.

renew: A dummy variable equal to 1 if pro-renewable-energy policies existed in destination country j in year t–1. I lag this variable by one year because these policies often do not begin having effects immediately. It is important to consider whether any policies in the destination country regarding renewable energy may also attract solar-technology inflows. A wide range of policies can be used to encourage alternative-energy R&D and production, including feed-in tariffs, subsidies, and tax incentives. I constructed this variable using the International Energy Agency's World Energy Outlook Policy Database. I expect the coefficient on this variable to be positive.

sun: A meteorological indicator that captures average hours of sunshine in country j per year. This data, from the World Meteorological Organization, was accessed via the United Nations Data Explorer. I expect a positive sign on *sun* since nations with more sunlight on average can be expected to produce more solar technology. However, there may be some ambiguity in this variable, particularly if production is being off-shored to a country with less sunlight. In addition, so many other factors determine production and use of solar technology, such as infrastructure and general economic performance, that meteorological data may capture only a small sliver of the solar-technology decision process.

6: Patent Equivalent Results

Almost 80 percent of green-energy patents are filed by six nations—Japan, the United States, Germany, France, the United Kingdom, and South Korea (Latif, Maskus, et al., 2011). Thus, it can be seen that technology in this sector is fairly concentrated at a national level. Figure 2 shows the top 20 destination countries for U.S. solar patents between 1952 and 2011. Japan, China, Canada, Australia, and Germany comprise the top five.

I examine the dataset trimmed to include only the years 1995 and earlier. This is to account for the fact that most patents receive 75 percent of all citations they will ever receive within 20 years of being filed. Thus, limiting the years in the analysis helps control for the problem that newer patents have fewer citations not necessarily because of lower quality, but because of their age. These data include 22[7] high-income nations. I define high-income as a per capita GDP of at least $20,000. I chose $20,000 because, with the exception of Mexico and Turkey, most OECD nations have per capita incomes no lower than around $20,000. Therefore, this level of GDP is less likely to result in the heterogeneity that would occur if I defined high-income as the World Bank does, at about $12,500 per capita GDP. In my dataset's high-income group, Luxembourg has the maximum per capita GDP, at US$101,450. Italy has the lowest, at US$29,847. The average per capita GDP of the high-income group is US$49,276 (World Bank, 2016).

The summary statistics and results can be seen in Tables 6-1 – 6-4. Looking at Table 6-1, we see that in an average year, the U.S. will file about 12 solar patents in country j. However, about 57 percent of all patents filed in the U.S. will not be filed elsewhere. About 31 percent of patents will be filed in two to 20 countries. The variable *qual* is quality measure of citations received by patents in this dataset weighted by the number of patents filed. Recall that this is calculated using an aggregate of the number of solar patents filed in each country each year. On average, the patent quality measure equals almost 4.5. Almost 61 percent of the p_{ijt} pairs receive no citations. The measure of IPR strength, *ipr*, has a mean of 2.6, above average in the Ginarte and Park index.

I now analyze the results of the negative binomial regression for the equivalents analysis. Table 6-2 shows the results without the sun variable. I elect to omit the sun variable in this iteration of the analysis because the correlation between *sun* and the dependent variable is very low, at 6.8 percent. Given the move toward green energy discussed in the introduction, I might expect policy, rather than hours of sunlight, to better predict the number of solar patents filed in a country. In this analysis, as expected, the coefficient on *qual* is positive; results indicate that the difference in logs of expected counts of p_{ijt} would decrease by approximately .12 units for a one-unit change in the aggregate quality ratio, while holding other variables in the model constant. In other words, a one-unit change in the aggregate quality ratio would cause p_{ijt} to decrease by about 12 percent. Using the summary statistics in Table 6-1, we can calculate the effect at the mean: A one-unit change in *qual* would lead to about 1.5 more patents being filed in a given year. We can also calculate the effect within one standard deviation, which would be almost 11 patents.

[7] The countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Norway, New Zealand, Singapore, Sweden, Switzerland, Taiwan, United Kingdom.

The coefficient on *ipr*, the other variable of interest, is positive as expected but not statistically significant. Looking at other independent variables in the analysis, we see that *gdp*, and *dist* are positive and negative, respectively, and statistically significant as predicted, indicating that richer economies and countries closer to each other will have more solar patents filed. Also as expected, the *lang* variable is negative, and it is statistically significant at the 5% level. This indicates that the increased cost of translation required in countries that do not share an official language with the U.S. may deter patent filings. The variable on *renew*, the dummy indicating whether a country has pro-renewable-energy policies in place, is negative but not statistically significant. It may be that my current proxy is not adequately picking up the effect that I want to measure. Alternately, there may be a feedback effect whereby higher innovation in renewable energies causes such policies to be created, not the other way around.

We can also examine the results including the variable *sun*, a meteorological measure of average hours of sunlight per day per country, in Table 6-3. In this case, *qual* is positive and statistically significant and *ipr* negative but not statistically significant. Results show that a one-unit change in the aggregate quality ratio would cause p_{ijt} to increase by about 9.3 percent. The effect at the mean translates to a rise of about 1.13 patents. The *sun* variable is positive and statistically significant at the 5% level; however, the actual effect is quite small: A one-hour increase in the yearly average amount of sunlight a country experiences would lead to .0046% more patents being filed in a given year. In this specification, the variable representing the need for translation, *lang*, is statistically significant, indicating that not sharing an official language with the U.S. decreases the number of patents filed at the mean by almost 9 per year. As in Table 6-2, *renew* is again negative and statistically insignificant.

A scatterplot of *sun* against the dependent variable, p_{ijt}, reveals a nonlinear relationship. It may be that hours of sunlight or quality alone do not affect the propensity to patent solar technology in a country, but that combined, these two have an effect. Therefore, I interact the *sun* variable with *qual*, the weighted aggregate quality measure. These results can be seen in Table 6-4. As with the results in Table 6-3, the weighted aggregate measure of quality, *qual*, is positive and statistically significant, while *ipr* is negative and statistically insignificant. The *sun_qual* interaction term is negative and statistically insignificant; however, *sun* is positive and statistically significant, albeit with a very small effect, as in Table 6-3. Looking at other independent variables for this analysis, results are similar to previous one: *lang* is statistically significant, while *renew* is not.

7: Extended Patent Families Results

The summary statistics and results can be seen in Tables 7-1 – 7-4. Looking at Table 7-1, we see that in an average year, the U.S. will file about 25 solar patents in country j. However, about 55 percent of all patents filed in the U.S. will not be filed elsewhere. About 22 percent of patents will be filed in two to 20 countries. The quality measure averages about 6.23. About 56 percent of the p_{ijt} pairs receive no citations. The measure of IPR strength, *ipr*, has a mean of 2.62, above average in the Ginarte and Park index.

Note the differences between the summary statistics in Tables 6-1 and 7-1. For the patent equivalent analysis, p_{ijt}, the average number of solar patents filed from the U.S. in country *j* in year *t* was about 12; for the extended families analysis, the average of p_{ijt} is about 25, a difference of more than 100 percent. The standard deviation of the p_{ijt} in the extended analysis is also larger by more than 100 percent: 25 compared with 52. The average quality measure for equivalents was about 4.5, while for extended families the average is a little more than 6. In this case, the standard deviation for the extended families analysis is much larger than that of the equivalents analysis (7 vs. 51). These differences are consistent with the definitions discussed in Section 3: Patent equivalents, or patent-to-patent analysis, provide a much narrower definition of what constitutes a patent flow, while extended patent families cover not just equivalents but also those patents that are both directly and indirectly related (recall Figure 1). Therefore, it is not surprising and indeed expected that in the extended families analysis, we would see higher average counts, leading to more patents to be cited, leading to a higher quality measure on average. Examination of the results for extended families reveals important differences as well.

Table 7-2 shows the results without the sun variable. Here, the coefficient on *qual* is negative and statistically significant at the 1 percent level, while for the equivalents analysis it was positive and statistically significant. These results indicate that the difference in logs of expected counts of p_{ijt} would decrease by approximately .004 units for a one-unit change in the aggregate quality ratio, while holding other variables in the model constant. This is an extremely small real effect: A one-unit change in the aggregate quality ratio would cause p_{ijt} to decrease by about 0.4 percent. Using the summary statistics in Table 7-1, we can calculate the effect at the mean: A one-unit change in *qual* would lead to about 0.1 fewer patents being filed in a given year. We can also calculate the effect within one standard deviation, which would be a bit more than 5 patents. This result indicates that lower-quality patents are more likely to be filed abroad, but only very slightly. This could be the policy effect mentioned previously—government regulations push the adoption of solar technology, regardless of quality. Or, since previous research has shown that patent quality and patent value are correlated, it is possible that lower-quality solar patents are more likely to be filed abroad when they are cheaper.

The coefficient on *ipr* is positive and statistically significant, while it was positive and statistically insignificant in the equivalent analysis. Results show that the difference in logs of expected counts of p_{ijt} would increase by about 1.18 units for a one-unit increase in the IPR index. In other words, a one-unit change in the IPR index would cause p_{ijt} to increase by about 118 percent. The effect at the mean translates to a rise of almost 30 patents. Ergo, this analysis supports the idea that stronger patent laws lead to an increase in the filing of patents abroad, rather than the lack of effect found in the equivalents analysis.

Looking at other independent variables in the analysis, we see that *gdp* is negative and statistically significant, while *imps* is positive and statistically significant as predicted. The results for *gdp* are a bit surprising, but the actual effect on this variable and *imps* is very small. The *lang* variable is not statistically significant, while the *renew* variable is statistically significant at the 10% level indicating that pro-renewable government policies may actually reduce patent inflows in the solar industry. This is a puzzling result. More study is needed, but it could be that government policies target and support established firms, which discourages patent flows. For example, some analysis has already occurred concerning the different effects that strengthened IPR laws could have on an importing country. Using aggregated data, Maskus and Penubarti (1995) found that stronger IPR laws may reduce imitation and encourage firms to increase exports to the country, thus causing a "market-expansion effect"; otherwise, such laws could reduce imitation and encourage firms to raise unit price, thus having a "market-power effect." The authors found that the market-expansion effect outweighed the market-power effect in data on OECD exports to the developing world, but it may be that effects differ when broken down by industry. In higher-income countries, it could be that "market-power" policies reduce patent flows.

We can now examine the results including the variable *sun*, in Table 7-3. In this case, *qual* is also negative and statistically significant, with a very small economic effect (0.08 fewer patents at the mean). *ipr* is positive and statistically significant. A one-unit change in the IPR index would cause p_{ijt} to increase at the mean by almost 19 patents. These results are the opposite of the equivalent analysis, where *ipr* was negative and statistically insignificant. The control variables perform similar to those in Table 7-2: *gdp* is negative and statistically significant, and *renew* negative and statistically insignificant at the 10% level. Compared with Table 7-2, however, *lang* becomes statistically significant at the 5% level when the *sun* variable is added, indicating that sharing a language with the origin country (U.S.), results in more solar patents being filed. The effect at the mean is fairly large, about 12 patents per year. *Sun* is also positive and statistically significant, but the economic significance is extremely small. One hour on average of more sunlight in country j translates to 0.02 more patents at the mean and about 10 more patents within one standard deviation.

Next, as discussed in the previous section, I interact the *sun* variable with *qual*. These results can be seen in Table 7-4. In this iteration, the weighted aggregate measure of quality, *qual*, is positive and statistically significant at the 5% level. The effect at the mean is about 2 patents; again, *ipr* is positive and statistically significant. A one-unit change in the quality measure would cause p_{ijt} to increase at the mean by about 19 patents. Compare this with the results of the equivalent analysis, where the coefficient on *ipr* is negative and statistically insignificant.

The *sun_qual* interaction term is negative and statistically significant, while it is negative and statistically insignificant in the equivalents analysis. However, both have a very small effect size. Looking at other independent variables for this analysis, *renew* and *lang* perform the same as they did in Table 7-3—*renew* is negative and statistically significant at the 10% level, while sharing a common language has a positive, statistically significant effect on the number of solar patents filed.

8: Conclusion

Table 8-1 summarizes the differences between the main variables of interest. It is striking that there is very little overlap, and this is the case for many of the other independent variables as well. If the question is: "Do strengthened IPR laws lead to more solar patenting abroad?" the answer depends on whether one examines patent equivalents or extended families. For the former, tightened IPR laws have no statistically significant effect, while for the latter, tightened IPR laws increase these patent flows. If the question is: "Are higher-quality solar patents more likely to be filed abroad?" the answer again depends on equivalents vs. extended families. Quality has a positive effect in all equivalents analyses; however, it has a negative effect in the families analysis unless the sun interaction term is considered. What can explain these divergent results? Which ones are more accurate? Solving these issues is key to crafting good intellectual-property policy. If strengthening IPR laws reduces solar technology transfer, this is an important consideration for policymakers and governments that are trying to expand green energy. Likewise, if strong IPR laws encourage the transfer of solar technology, then policymakers have a tool to use to achieve this goal.

It is clear that the way patent equivalents vs. extended patent families are defined affects significantly how they are counted, as seen in the comparison of summary statistics for both analyses. One might expect that because the extended family analysis has more patents being filed on average, and a larger average quality measure than the equivalents analysis, that the quality variable would yield a statistically significant result in this group. However, a look inside the extended families data may reveal why this is not the case. Although a formal data analysis has not been conducted, I was able to complete two small case studies of patents not included in this data analysis. These examine solar patents from the U.S. to Greece in the year 1992 and to Malaysia in 1995. (These case studies were chosen arbitrarily from a larger, but less complete in terms of control variables, dataset.) The Greece 1992 case has data from four patent families. The first family contains 76 documents, but about 90 percent of the citations come from only 15 of those documents. The second family contains 12 documents, but about 80 percent of the citations come from only three of those patents. The third family contains 11 patents, but 90 percent of the citations come from only one of those patents. The fourth family contains 12 documents, but 95 percent of the citations come from one document. The Malaysia 1995 study reveals a different issue: In this case, there are six families in the data that year, but over 95% of the citations come from one family.

What these case studies reveal is that extended families data contain a significant amount of "noise" that most likely affects the results. In other words, only a few patents in families, or only a few families, account for the large majority of citations, so the quality measure (which is a ratio of citations/patents filed) is in reality much smaller, even though the total number of citations and patents filed is, on average, much larger. This "noise" can also be seen in the summary statistics. Note that the maximum value of the aggregate weighted quality measure for the extended families analysis is 1487, while it is 68.5 for the equivalents analysis. Perhaps outliers (a few patents that receive a very large number of citations) are biasing the results. To try to answer this question, I re-ran the analysis, dropping observations where *qual* exceeded 100. Although not reported in detail here, the results stayed the same in terms of sign, coefficient size, and significance. The main difference is that the average of p_{ijt} increased by about 70 percent, meaning that effect sizes at the mean would be even larger in this case. Therefore, even controlling for some of the "noise" doesn't seem to be able to resolve the different results obtained with the different data.

However, the equivalents method may not be any better. As noted in Section 3, if the intent is to track the transfer of complete technologies, the equivalent method could result in over-counting patents and therefore inflating citations and results. The only way to determine if that is the case is to compare the individual patents contained in each dataset. That is beyond the scope of this study, but it is a direction for future research.

Other methods have also been proposed to solve the issue of patent equivalents vs. patent families. First, Martínez (2011) describes using "expert-validated families based on novel technical content" to determine whether patents actually belong in a particular family and are directly related. This solves the problem of families being defined differently across patent offices and organization while also ensuring that patents contained in the families do indeed cover the same invention. She cites two examples: the EPO's DOCDB families database and the Derwent World Patent Index (DWPI). Conducting a similar analysis using these databases could help resolve the different results found here.

Second, using firm-level data would also allow researchers to follow technology flows more specifically and closely. However, accessing firm-level data, particularly when IPRs are involved, can be difficult. Although patents are publicly available, not all list firms as inventors, so the quality of data culled from these sources might be questionable.

This research initially set out to discover whether quality and IPR laws are important determinants of patent flows and technology transfer in the solar sector. The answer depends both upon which method of patent analysis is used. Although this paper cannot resolve the question, it is of vital importance to current and future policies aimed at increasing access to climate-change-mitigating technologies. Moreover, the results presented here should give patent researchers pause, and encourage us all to consider more carefully how we define and use our data.

Figure 1:

Illustration of Patent Equivalent vs. Extended Patent Family

In this case, document D1 is the only document in family P1, D2 and D3 belong to family P1-P2, D4 belongs to family P2-P3, and D5 to family P3.

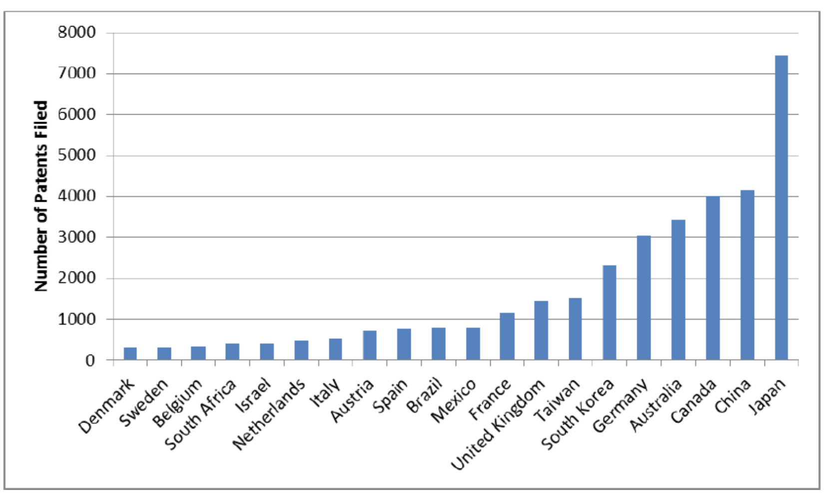
Document D1	Priority P1			FAMILY P1
Document D2	Priority P1	Priority P2		FAMILY P1-P2
Document D3	Priority P1	Priority P2		FAMILY P1-P2
Document D4		Priority P2	Priority P3	FAMILY P2- P3
Document D5			Priority P3	FAMILY P3

If all the priorities of two documents are the same, they are referred to as "equivalents". This definition is currently used in Espacenet for listing the documents under "also published as" on the bibliographic data view.

Source: European Patent Office, "Definitions," Accessed 10-14-2016.

Figure 2:

Top 20 destination countries for U.S. solar patents, 1952–2011



Source: Dataset used for this paper.

Table 6-1:

Equivalent Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
p_{ijt}	880	12.17159	25.10576	0	141
qual	880	4.484949	7.20212	0	68.5
ipr	704	2.601391	0.6738054	1.257353	4.341667
humk	880	0.2544807	0.1855406	0.0295	1.0572
gdp	846	147897.5	314904.9	157.7524	2802013
dist	880	8239.227	3768.445	733	15943
imps	630	4700652	9952884	10607	76100000
renew	588	0.0238095	0.1525851	0	1
lang	880	0.6818182	0.4660354	0	1
sun	720	1847.8	509.4229	1157.1	3353.55

Table 6-2:

Equivalent Results w/o Sun

Variable	Coeff./SE
constant	.8924427
	[0.5732192]
qual	0.1231126***
	[0.0172176]
ipr	0.2824085
	[0.1886843]
humk	1.252196**
	[0.5109865]
gdp	0.000023***
	[0.0000037]
dist	-0.0000793***
	[0.0000279]
imps	-0.0000000484
	[0.0000000101]
renew	-0.4692364
	[0.4067666]
lang	-0.4511966**
	[0.1845646]
N	488
X^2	235.14

Standard errors in brackets

*** = statistically significant at 1% level; ** = statistically significant at 5% level; * = statistically significant at 10% level

Table 6-3:

Equivalent Results w/ Sun Only

Variable	Coeff./SE
constant	1.702015**
	[0.6762208]
qual	0.0931931***
	[0.0168053]
ipr	-0.1077542
	[0.1969152]
humk	0.8180235
	[0.5893785]
gdp	0.0000455***
	[0.00000805]
dist	-0.0000626
	[0.0000437]
imps	-0.00000153***
	[0.000000469]
renew	-0.3449747
	[0.3853938]
lang	-0.7161813***
	[0.2032459]
sun	0.0004577**
	[0.000187]
N	376
X^2	171.48

Table 6-4:

Equivalent Results w/ Sun-Qual Interaction Term

Variable	Coeff./SE
constant	1.293302*
	[0.7599621]
qual	0.1525595***
	[0.0553339]
ipr	-0.1226899
	[0.1960662]
humk	0.8722276
	[0.5909239]
gdp	0.0000451***
	[0.00000804]
dist	-0.0000646
	[0.0000438]
imps	-0.000000153***
	[0.000000468]
renew	-0.353933
	[0.3855548]
lang	-0.6996096***
	[0.2042011]
sun	0.000681**
	[0.0002819]
sun_qual	0.0000303
	[0.0000265]
N	376
X^2	172.72

Table 7-1:

Extended Families Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
p_{ijt}	1012	25.50395	52.22216	0	355
qual	1012	6.226591	51.4575	0	1487
ipr	792	2.626641	0.7693044	1.25	4.54
humk	1012	0.2711067	0.2049988	0.03	1.09
gdp	991	15108.09	7071.82	1555.43	46141.95
dist	1012	8239.227	3768.165	733	15943
imps	672	6110000000	1290000000	10600000	10900000000
renew	1012	0.0395257	0.1949384	0	1
lang	1012	0.3181818	0.4660008	0	1
sun	828	1847.798	509.3767	1157.1	3353.55

Table 7-2:

Extended Families Results w/o Sun

Variable	Coeff./SE
constant	-0.011233
	[0.4370754]
qual	-0.0039133***
	[0.0012952]
ipr	1.182623***
	[0.145328]
humk	1.032853**
	[0.4188254]
gdp	-0.0000529***
	[0.0000191]
dist	0.0000385**
	[0.0000193]
imps	0.000000000431***
	[0.00000000008]
renew	-0.4476046*
	[0.2650107]
lang	-0.066827
	[0.1504678]
N	672
X^2	234.57

Standard errors in brackets
*** = statistically significant at 1% level;
** = statistically significant at 5% level;
* = statistically significant at 10% level

Table 7-3:

Extended Families Results w/ Sun Only

Variable	Coeff./SE
constant	.5272198
	[0.5217214]
qual	-0.0032577***
	[0.0011456]
ipr	0.7430126***
	[0.1659381]
humk	1.372051**
	[0.5538578]
gdp	-0.0000455**
	[0.0000209]
dist	-0.000126***
	[.0000368]
imps	0.000000000945***
	[0.000000000142]
renew	-0.5050316*
	[0.2813885]
lang	0.4661196**
	[0.1852984]
sun	0.0008013***
	[0.0001967]
N	573
X^2	227.89

Table 7-4:

Extended Families Results w/ Sun-Qual Interaction Term

Variable	Coeff./SE
constant	0.362887
	[0.525267]
qual	0.0816531**
	[.0329503]
ipr	0.7450482***
	[0.1660626]
humk	1.161476**
	[0.5553454]
gdp	-0.0000441**
	[.0000209]
dist	-0.0001193***
	[0.0000365]
imps	0.000000000952***
	[0.000000000142]
renew	-0.4843313*
	[0.2802114]
lang	0.5063027***
	[0.1852048]
sun	0.000873***
	[0.0001959]
sun_qual	-0.0000483***
	[0.0000187]
N	573
X^2	233.54

Table 8-1:

Summary of Results/Differences

	Equivalent Analysis	Extended Families Analysis
w/o sun	qual: + / SS ipr: + / NSS	qual: – / SS ipr: + / SS
w/sun only	qual: + / SS ipr: – / NSS	qual: – / SS ipr: + / SS
w/sun_qual interax	qual: + / SS ipr: – / NSS sun_qual: – / NSS	qual: + / SS ipr: + / SS sun_qual: – / SS

SS = Statistically significant

NSS = Not statistically significant

Appendix 1:

Data-Collection Process for Patent Equivalent Method

Below is a step-by-step list of the process used to gather data on the patents and citations used in this study.

1. The program read the list of IPC codes
2. For given year and month range, the program

 a. Accessed web pages using the following templates

http://worldwide.espacenet.com/searchResults?page=0&IC=[CODE]&DB=EPODOC&PD=[YEAR][MONTH]&locale=en_EP&ST=advanced&compact=false

http://worldwide.espacenet.com/searchResults?page=0&IC=H01L31/00&DB=EPODOC&PD=197001&locale=en_EP&ST=advanced&compact=false

 b. Downloaded the HTML code and stored it on the local hard disk

 c. If the file did not contain all the results, the search was narrowed and broken down into days. In that case, the program downloaded HTML using the following template

http://worldwide.espacenet.com/searchResults?page=0&IC=[CODE]&DB=EPODOC&PD=[YEAR][MONTH][DAY]&locale=en_EP&ST=advanced&compact=false

 d. If in one day too many patents were filed for the results to be returned on a single web page, corresponding files were downloaded manually

3. For each search result HTML file, the following actions were performed

 a. Files with more than 15 search results were ignored, as these files were broken down into smaller files as described above in 2

 b. Each link in the search results were followed and its HTML content downloaded

 c. Each downloaded file was analyzed to search for a “more” button, which linked to additional information. If the phrase was present, the program followed it and download complete data

 d. The link “View list of citing documents” was followed and its HTML content downloaded (This option may no longer be available on the website)

4. For each downloaded citations file, the program then

a. Checked if the file had a reference to a “next” button (that is it, checked if the citations were listed on more than one page)

b. If needed, downloaded the file with the next portion of citations

c. Redid parts a. and b. until all citations were downloaded

5. The program created a CSV data file “patents.csv” and

a. Loaded each stored patent file and read and extracted the data from the file

b. Appended the data into corresponding columns of the CSV file

6. Then the program created a CSV data file “citations.csv” and

a. Loaded each stored citation file and read and extracted the data from the file

b. Appended the data into corresponding columns of the CSV file

7. At each step, the program checked for the integrity of the downloaded data by ensuring that the entire file was downloaded, that it really contained data rather than an error message, etc. After this procedure, the files were ready to be imported into data-analyzing software (MS Access, Stata). The specific code used in this process is available upon request, as are the downloaded files themselves.

Appendix 2:

Data-Collection Process for Extended Patent Families Method

The data about patents was obtained from EPO using their OPS EPO services.

1. The first step in the data collection process was to obtain a list of patent IDs with the chosen IPC codes. The number of search results was limited to 100, and most of the IPC codes had the number of patents vastly exceeding 100. Therefore, the query for each IPC code was split using date intervals into a number of queries yielding at most 100 results each. As a result, 10,193 XML files containing search results were downloaded from OPS EPO services.

2. These files were subsequently parsed in order to extract patent IDs. A list of 546,029 patent IDs was created.

3. Then, for each patent, OPS EPO services were queried about their patent family information as well as the citation count. A total of 1,092,058 XML files with bibliographic information were downloaded (out of which three were corrupt; attempts to manually access EPO database resulted in database failure).

4. These files were subsequently parsed for the information about their priority numbers, numbers of citations, and family members. The parsed data was stored in CSV files and later uploaded to Microsoft Access for further analysis. The citation numbers were combined for families, and then each patent was reassigned a citation number corresponding to its family. Each patent was also assigned a priority country (source country), application country (destination country), and application date (year).

5. Finally, the data about patents were aggregated into source-destination-year units on which the regressions were performed.

References:

Acosta, Daniel, Daniel Coronado, and Ana Fernández, “Exploring the quality of environmental technology in Europe: evidence from patent citations,”*Scientometrics*, 80, no. 1 (2009), 133–154.

Adams, Samuel, “Intellectual Property Rights, Investment Climate and FDI in Developing Countries” *International Business Research*, Vol. 3, No. 3, July 2010.

Adams, Steven, *Information Sources in Patents*, 3rd Ed. (Berlin/Boston: Walter de Gruyter GmgH & Co. KG, 2012), p. 15.

Barro, Robert J., and Jong Wha Lee. "A new data set of educational attainment in the world, 1950–2010." *Journal of development economics* 104 (2013): 184-198.

Barton, John, "Intellectual Property and Access to Clean Energy Technologies in Developing Countries: An Analysis of Solar Photovoltaic, Biofuel and Wind Technologies," ICTSD Issue Paper No. 2, 2007.

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

Branstetter, Lee, Raymond Fisman, and C. Fritz Foley, "Do Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from U.S. Firm-Level Panel Data," *The Quarterly Journal of Economics*, 121, no. 1, (2006), 321–349.

Branstetter, Lee, Raymond Fisman, C. Fritz Foley, and Kamal Saggi, "Intellectual Property Rights, Imitation, and Foreign Direct Investment: Theory and Evidence," NBER Working Paper No. 13033, April 2007.

Chen, Yu-Shan and Ke-Chiun Chang, "The Relationship Between a Firm's Patent Quality and Its Market Value—The Case of the U.S. Pharmaceutical Industry," *Technological Forecasting and Social Change*, 77, no. 1, (2010), 20–33.

CIA World Factbook.

Dechezleprêtre, Antoine; Martin, Ralf; and Mohnen, Myra, "Knowledge Spillovers from Clean and Dirty Technologies: A Patent Citation Analysis," Working Paper, September 29, 2013.

Dechezleprêtre, Antoine; Glachant, Matthieu; Haščič, Ivan; Johnstone, Nick, and Ménière, Yann, "Invention and Transfer of Climate Change–Mitigation Technologies: A Global Analysis," *Review of Environmental Economics and Policy*, Vol. 5, Issue 1, Winter 2011, pp. 109–130.

Dechezleprêtre, Antoine, Matthieu Glachant, and Yann Ménière. "What drives the international transfer of climate change mitigation technologies? Empirical evidence from patent data." *Environmental and Resource Economics* 54.2 (2013): 161-178.

De Rassenfosse, Gaetan and van Pottelsberghe de la Potterie, Bruno, "The Role of Fees in patent Systems: Theory and Evidence," *Journal of Economic Surveys*, Volume 27, Issue 4, September 2013, 696–716.

De Rassenfosse, Gaetan and van Pottelsberghe de la Potterie, Bruno, "On the Price Elasticity of Demand for Patents," *Oxford Bulletin of Economics and Statistics*, Vol. 74, Issue 1, p.p. 58–77, 2012.

Eaton, Jonathan and Kortum, Samuel, "International Patenting and Technology Diffusion," NBER Working Paper No. 4931, November 1994).

Eaton, Jonathan, and Samuel Kortum. "Trade in ideas Patenting and productivity in the OECD." *Journal of international Economics* 40.3 (1996): 251-278.

Espacenet, "Patent Families," Accessed 10-14-2016.

European Commission, "Proposal for a Council Regulation on the Translation Arrangements for the European Union Patent," 2010.

European Patent Office Espacenet Patent Database.

European Patent Office, Definitions, Accessed 10-14-2016.

Evenson, Robert and Kanwar, Sunil, "On the Strength of Intellectual Property Protection That Nations Provide," *Journal of Development Economics*, 90 (2009) 50–56.

Evenson, Robert, and Kanwar, Sunil, "Does Intellectual Property Protection Spur Technological Change?" Yale University Economic Growth Center Discussion Paper No. 831, June 2001.

Fallah, M. Hosein, Elliot Fishman, and Richard R. Reilly, "Forward Patent Citations as Predictive Measures for Diffusion of Emerging Technologies," PICMET Proceedings, August 2–6, 2009.

Feenstra, Robert C., and Robert E. Lipsey, NBER–United Nations Trade Data, 1962 2000.

Fitzpatrick, Gary L. and Modlin, Marilyn J., *Direct-Line Distances: International Edition* (Scarecrow Press, 1986).

Gallini, Nancy; Putnam, Jonathan; and Tepperman, Andrew, "Intellectual Property Rights and the Propensity to Patent," in J. Putnam (ed.), *Intellectual Property Rights and Innovation in the Knowledge-Based Economy*, (Ottawa: Industry Canada, 2006).

Ginarte, Juan and Park, Walter, "Determinants of Patent Rights: A Cross-National Study," *Research Policy* 26 (1997) 283-301.

Haščič, I. and M. Migotto (2015), "Measuring environmental innovation using patent data," *OECD Environment Working Papers*, No. 89, OECD Publishing, Paris.

Hall, Bronwyn and Helmers, Christian, "The Role of Patent Protection in Clean/Green Technology Transfer," NBER Working Paper No. 16323, September 2010.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, "Market Value and Patent Citations," *RAND Journal of Economics*, 36, no. 1 (2005), 16–38.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, "The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools," NBER Working Paper No. 8498, October 2001.

Harhoff, Dietmar, "Measuring and estimating patent value," WIPO-OECD Workshop on Statistics in the Patent Field, 2003.

Harhoff, Dietmar, Francis Narin, F.M. Scherer, and Katrin Vopel, "Citation Frequency and the Value of Patented Inventions," *The Review of Economics and Statistics*, 81, no. 3 (1999), 511–515.

Heston, Alan; Summers, Robert; and Aten, Bettina, Penn World Table Version 7.0, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, May 2011.

Hirschey, Mark, and Vernon Richardson, "Valuation Effects of Patent Quality: A Comparison for Japanese and U.S. Firms," *Pacific Basin Finance Journal*, 9, no. 1 (2001), 65–82.

Intellogist, "Patent Families," accessed 10-14-2016.

International Energy Agency RD&D Statistics.

International Monetary Fund, World Economic Outlooks Database.

Jaffe, Adam B. and de Rassenfosse, Gaétan, “Patent Citation Data in Social Sciences Research: Overview and Best Practices,” NBER Working Paper 21868, January 2016.

Javorcik, Beata Smarzynska, “The Composition of Foreign Direct Investment and Protection of Intellectual Property Rights: Evidence from Transition Economies,” *European Economic Review*, 48 (2004), 39–62.

Kanwar, Sunil, “Intellectual Property Protection and Technology Transfer: The Case of Overseas R&D,” Delhi School of Economics Centre for Development Economics Working Paper No. 166, April 2009.

Kortum, Samuel, and Josh Lerner. "Stronger protection or technological revolution: what is behind the recent surge in patenting?" *Carnegie-Rochester Conference Series on Public Policy*. Vol. 48. North-Holland, 1998.

Kumar, Nagesh, “Intellectual Property Protection, Market Orientation and Location of Overseas R&D Activities by Multinational Enterprises,” *World Development*, 24, no. 4 (1996), 673–688.

Lall, Sanjaya, “Indicators of the Relative Importance of IPRs in Developing Countries,” ICTSD–UNCTAD Issue Paper No. 3, 2003.

Lanjouw, J.O., and M. Schankerman, “The Quality of Ideas: Measuring Innovation with Multiple Indicators,” NBER Working Paper No. 7345, September 1999.

Lanjouw, “Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators,” *Economic Journal*, 114, no. 495 (2004), 441–465.

Latif, Ahmed Abdel; Maskus, Keith; Okediji, Ruth; Reichman, Jerome; and Roffe, Pedro, “Overcoming the Impasse on Intellectual Property and Climate Change at the UNFCCC: A Way Forward,” International Centre for Trade and Sustainable Development, Policy Brief 11, November 2011.

Mailänder, Lutz, “Patent Families,” WIPO Presentation, December 10, 2012.

Marco, Alan C., “The Dynamics of Patent Citations,” *Economics Letters*, 94, no. 2 (2007), 290–296.

Martínez, Carolina. “Patent Families: When Do Definitions Really Matter?” *Scientometrics* (2011) 86:39–63.

Maskus, Keith, “Encouraging International Technology Transfer,” ICTSD–UNCTAD Issue Paper No. 7, May 2004.

Maskus, K., and L. Yang, “Intellectual Property Rights and Licensing: An Econometric Investigation,” in *Intellectual Property and Development: Lessons from Recent Economic Research*, C. Fink, and K. Maskus, eds. (Washington, DC: The International Bank for Reconstruction and Development/The World Bank, 2005).

Maskus, Keith, “The Role of Intellectual Property Rights in Encouraging Foreign Direct Investment and Technology Transfer,” in *Intellectual Property and Development: Lessons from Recent Economic*

Research, C. Fink, and K. Maskus, eds. (Washington, DC: The International Bank for Reconstruction and Development/The World Bank, 2005).

Maskus, Keith, "Patents and Technology Transfer Through Trade and the Role of Regional Trade Agreements," Think-piece prepared for MCTI Conference, East-West Center, January 2016.

Maskus, Keith, and Mohan Penubarti, "How Trade-Related Are Intellectual Property Rights?" *Journal of International Economics*, 39, (1995), 227–248.

Meléndez-Ortiz, Ricardo, "Enabling the Energy Transition and Scale-up of Clean Energy Technologies: Options for the Global Trade System," E-15 Initiative, International Centre for Trade and Sustainable Development and World Economic Forum, January 2016.

Norback, Pehr-Johan, Lars Persson, and Roger Svensson, "Creative Destruction and Productive Preemption," Center for Economic Policy Research Discussion Paper Series No. 8281, March 2011.

Phalin, Amanda (listed as Watson, Amanda, maiden name), "Does TRIPS Increase Technology Transfer to the Developing World? The Empirical Evidence," *Information and Communications Technology Law*, Vol. 20, No. 3, October 2011, 253–278.

Phalin, Amanda, "International Transfer of Solar Technology," Dissertation, University of Florida, 2013.

Pillu, Hugo, and Gilles Koléda, "Induced Innovation and International Technological Opportunity in the Field of Energy: Evidence from World Patent Citations," Equipe de Recherche en Analyse des Systèmes de Mondialisation Economique Paper,June 2009.

Popp, "International Technology Transfer, Climate Change, and the Clean Development Mechanism," *Review of Environmental Economics and Policy*, 5, no. 1 (2011), 131–152.

Popp, David, Ivan Haščič, and Neelaksi Medhi, "Technology and the Diffusion of Renewable Energy," *Energy Economics*, 33, (2011), 648–662.

Popp, D., "They Don't Invent Them Like They Used To: An Examination of Energy Patent Citations Over Time," *Economics of Innovation and New Technology*, 15, no. 8 (2006), 753–776.

Rafiquzzaman, Mohammed and Whewell, Lori, "Recent Jumps in Patenting Activities: Comparative Innovative Performance of Major Industrial Countries, Patterns, and Explanations," Industry Canada Research Publications Program Working Paper No. 27, December 1998.

REN21, "Renewables 2016 Global Status Report," Renewable Energy Policy Network, 2016.

Rosenkopf, Lori, and Atul Nerkar, "Beyond Local Search: Boundary-Spanning, Exploration, and Impact in the Optical Disk Industry," *Strategic Management Journal*, 22, (2001), 287–306.

Scherer, Frederic, "Using Linked Patent and R&D Data to Measure Interindustry Technology Flows," in *R&D, Patents, and Productivity*, Zvi Griliches, ed. (Chicago: University of Chicago Press, 1984).

Trajtenberg, Manuel, "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *The RAND Journal of Economics*, 21, no. 1 (1990), 172–187.

United Nations Data Explorer.

United Nations Comtrade Database.

Van Zeebroeck, Nicolas, "The puzzle of patent value indicators," *Economics of Innovation and New Technology*, 20, no. 1 (2011), 33–62.

World Bank, "GDP Per Capita," 2016, accessed 10-25-2016.

World Resources Institute (WRI), "Statement: A Climate Deal Comes Together in Durban," Accessed online 2-23-2012.

World Resources Institute (WRI), "The Paris Agreement: Turning Point for a Climate Solution," Accessed online 1-30-2016.