OECD

**USING THE TRIADIC PATENT FAMILY DATABASE TO STUDY
ENVIRONMENTAL INNOVATION**

This report was prepared by Dr. David Popp (Department of Public Administration, Maxwell School, Syracuse University). The intention is to assess the potential for the use of the OECD's Triadic Patent Family Database in the analysis of the links between environmental policy and technological innovation, and to provide suggestions for research methods in this area.

Abstract

Interest in the links between environmental policy and technological change has grown in recent years. One reason for this growth is the increased availability of detailed patent statistics, which can be used to provide a measure of technological change within the field of environmental research. This report discusses the advantages and disadvantages of using such patent indicators, with a particular focus on the newly developed OECD Triadic Patent Family (TPF) database. The database includes patent applications filed at the European Patent Office (EPO), United States Patent and Trademark Office (USPTO), and the Japanese Patent Office (JPO). International comparisons of patent data can be troubled by differences in standards between patent offices. In contrast, because the TPF includes only patents that have at least one family member in each of these three offices, the patents included in the database have all met a common threshold. This report discusses the pros and cons of using the TPF database to examine the links between environmental policy and innovation, and proposes methodology that should be of use to researchers using this database. Preliminary results suggest that, while the database can provide some useful information for evaluating environmental technologies across countries, it cannot serve as a stand-alone database for such research. Because of the limitations of the Triadic Patent Family database, other, more general patent databases should also be used.

Using the Triadic Patent Family Database to Study Environmental Innovation

David Popp

Department of Public Administration
The Maxwell School
Syracuse University

and

Faculty Research Fellow
National Bureau of Economic Research

I. Using Patents as an Indicator of Technological Change[1]

Several possible data sources exist for studying technological change and the environment. Research and development (R&D) data offer a straightforward measure of innovative activity. Diffusion studies often make use of adoption data for specific technologies. In comparison, patent data are not as straightforward. Patent counts should be expected to increase as R&D activity increases, but the correlation need not be exact. Furthermore, the existence of a patent does not mean that the technology has been adopted. Indeed, studies of the economic value of patents find that most patents have little commercial value, suggesting that adoption of most patented inventions is not widespread (see, for example, Lanjouw *et al*. 1998). Nonetheless, patent data offer several advantages when studying technological change and its effect on the environment.

Unlike more aggregate data such as R&D expenditures, patents provide a detailed record of each invention. From the bibliographic data on a patent, the researcher can learn the identity and home country of the inventor, read a description of the invention, and see references to earlier patents. Using patent data, it is possible for researchers to collect data in highly disaggregated forms. Whereas R&D data are typically available only for specific industries or general applications,[2] patent classifications can be used to distinguish between different types of R&D at great detail, such as air pollution control devices designed to reduce NO_X emissions versus devices designed to control SO_2 emissions.[3] In addition, economists have found that patents, sorted by their date of application, provide a good indicator of R&D activity, as patent applications are usually filed early in the research process (see, for example, Griliches

---

[1] This section borrows heavily from Popp (forthcoming).

[2] For example, in the US, R&D data is available from 1972-1994 for air pollution control, but it is not broken down by pollutant.

[3] For example, US patent classes 423/235-423/239 pertain to control of "nitrogen or (a) nitrogenous component", and patent classes 423/242 – 423/244 and 423/569 – 423/570 pertain to control of sulfur compounds. Using patent databases, it is possible to download all patents in these classes.

1990). As a result, patent counts not only serve as a measure of innovative output, but are indicative of the level of innovative activity itself.

Furthermore, patent data are available from many different countries. Sources such as the European Patent Office make available data from virtually every country.[4] Of course, such data can be used to examine levels of innovative activity across countries. In addition, unlike R&D data, international patent data can be used to track patterns of diffusion. Because the legal protections granted by a patent only apply in the country in which the patent has been granted, inventors must file patent applications in each country for which they desire protection. As this provides a signal that the inventor views each county as a potential market, researchers can use such data to track diffusion of knowledge across countries (e.g. Lanjouw and Mody 1996).

Despite these advantages, for properly designed research, it is important to be aware of the limitations of patent data. The existing literature on the benefits and drawbacks of using patent data is quite large.[5] Most importantly, the quality of individual patents varies widely. Some inventions are extremely valuable, whereas others are of almost no commercial value. This is partly a result of the random nature of the inventive process. Accordingly, the results of studies using patent data are best interpreted as the effect of an "average" patent, rather than any specific invention. Alternatively, other information about the patent, such as subsequent patent citations or the family size, can be used to weight individual patents and control for quality differences (Lanjouw *et al.*, 1998).[6]

Second, Popp (2003) notes that the effect of a patent may change as policy changes. This work examines the effect of the U.S. SO_2 permit trading market on innovation for flue gas desulfurization (FGD). Prior to the SO_2 market, firms faced a best available control technology regulation. In response, there was much innovation, but this innovation focused on reducing the operating costs of the mandated FGD units. It was not until the more flexible permit market was put in place that the nature of R&D shifted towards developing FGD units that removed a larger percentage of SO_2 from the emission stream. A simple count of FGD patents would not reveal this change in the *nature* of innovation on FGD units. Thus, while patent counts provide a good measure of the *level* of innovative activity, they must be combined with other information if one desires to understand the *effect* of technological change.[7]

Finally, although the decision to file a patent obviously follows from the decision to perform R&D, not all successful research results are patented. In return for receiving the monopoly rights inferred by a patent, the inventor is required to publicly disclose the invention. Rather than make this disclosure, inventors may prefer to keep an invention secret. Surveys of inventors indicate that the rate at which new innovations are patented varies across industry (Levin *et al.* 1987). As such, this propensity must be controlled for. Methodology for controlling for the propensity to patent is discussed in Section V.

---

[4] Details about the data available through the esp@cnet, the European Patent Office's interface for world patent data, can be found at http://ep.espacenet.com/.

[5] Griliches (1990) provides a useful survey.

[6] Moreover, Popp (2003) notes that the effect of a patent may change as policy changes. This work examines the effect of the U.S. SO_2 permit trading market on innovation for flue gas desulfurization (FGD). Prior to the SO_2 market, firms faced a best available control technology regulation. In response, there was much innovation, but this innovation focused on reducing the operating costs of the mandated FGD units. It was not until the more flexible permit market was put in place that the nature of R&D shifted towards developing FGD units that removed a larger percentage of SO_2 from the emission stream. A simple count of FGD patents would not reveal this change in the *nature* of innovation on FGD units.

[7] However, this shortcoming is not unique to patents. Using R&D data to measure innovative activity would suffer from the same problems.

Fortunately, when studying the development of a single technology, this last problem is less of a concern than when using patent data to measure innovation trends across several dissimilar industries. However, it does have one important implication for environmental research. In general, firms are more likely to patent research that results in new *products*, rather than research that results in new *processes*. Because new products will be publicly available in the market, the loss of secrecy that comes with a patent is less of a concern than for an innovative process. Thus, firms will often choose to keep new process innovations secret, rather than patent them. As such, research on environmental innovation that uses patent statistics is likely to focus primarily on end-of-the-pipe treatment of pollution, rather than changes to the production process that occur in response to environmental regulation.

II. Understanding the Triadic Patent Family Database

As noted above, patents are granted by national patent offices in individual countries. Because patent protection is only valid in the country that grants the patent, an inventor must file for protection in each nation in which protection is desired. Nearly all patent applications are first filed in the home country of the inventor. The date of the initial application is referred to as the *priority date*. If the patent is granted, protection begins from the priority date. Inventors who desire patent protection in other nations must file applications in those nations, either directly or by using a Patent Cooperation Treaty (PCT) that designates the countries in which protection is desired, within one year of the priority date. If the inventor does file abroad within one year, the inventor will have priority over any patent applications received in those countries since the priority date that describe similar inventions.

Additionally, in Europe, inventors may choose to file an application through the EPO, rather than applying to individual patent offices. The applicant designates as many of the 18 EPO member-states for protection as desired. The application is examined by the EPO. If granted, the patent is transferred to the individual national patent offices designated for protection. Because EPO applications are more expensive, European inventors typically first file a patent application in their home country, and then apply to the EPO if they desire protection in multiple European countries.

These additional filings of the same patent application (which share the same priority date) in different countries are known as *patent families*. Because of the additional costs of filing abroad, along with the one-year waiting period that gives inventors additional time to gauge the value of their invention, only the most valuable inventions are filed in several countries. Moreover, filing a patent application in a given country is a signal that the inventor expects the invention to be profitable *in that country*. Because of this, researchers such as Lanjouw and Schankerman (2004) have used data on patent families as proxies for the quality of individual patents.

The TPF database makes use of this relationship between patent quality and patent families to provide a database of "high-quality" inventions. This offers a couple of advantages, particularly for macro-level evaluations of technological change. Unfortunately, as I'll discuss later, it also poses some challenges for evaluating environmental innovation. The first advantage is that the use of patent families limits the data to 'high quality' patents. There is much variation in the quality of inventions represented by individual patents. Some inventions are extremely valuable, whereas others are of almost no commercial value.[8] However, because of the added cost of filing abroad, the least valuable patents are usually filed only in the inventor's home country. In contrast, choosing to file for protection in the EPO (and thus designating several European countries for protection), the United States, and Japan suggests that the inventor anticipates his patent to be of greater value. As such, the TPF will contain fewer patents of little practical value than would patent counts taken from a single patent office.

---

[8] See, for example, Schankerman and Pakes (1986).

Second, as noted by Dernis and Kahn (2004), patent counts taken from individual patent offices include a home country bias. Because inventors almost always file for protection in their home country first, the majority of patents at the EPO, USPTO, and JPO come from domestic inventors. Thus, counts of patents from a single patent office do not provide an accurate comparison of the level of inventive activity across countries.[9] In contrast, since patent applications in the TPF have been filed at the EPO, USPTO, and JPO, the distribution of patents from each country should more carefully represent the distribution of inventive activity in these countries.[10]

A. Patent families in the OECD database

Patent families can be defined in several different ways. As a starting point, any patent application at the EPO, JPO, and patent grant at the USPTO sharing a common priority number may be considered a patent family. However, as noted in Dernis and Kahn (2004), this will overcount some innovations, due to the intricacies of individual patent offices. For example, patents at the JPO are typically narrower than patents taken at other offices – that is, each patent contains fewer claims than comparable patents abroad. This difference results from idiosyncrasies in Japanese patent law, which, until 1988, required a single patent application for each separate claim (Ordover 1991). For example, three separate priorities, representing three separate Japanese patents, might be shared by just one or two USPTO and EPO patents.

To account for this, the TPF database uses *consolidated* patent families. Any patents that are *directly* or *indirectly* linked to other patents are counted as a single patent family. For example, if patent A shares priorities 1 and 2, and patent B shares priorities 2 and 3, priorities 1 and 2 are directly linked, as are priorities 2 and 3. As they both share a direct link with priority 2, priorities 1 and 3 are *indirectly* linked. All patents sharing one of these three priorities are assembled in a single patent family in the OECD TPF database.

The advantage of such a methodology is that it results in counts that are comparable across countries. For example, using patent counts from individual patent offices, we expect to see a higher volume of patents from Japan, because of the need for multiple patents to represent a single invention. In addition, as a patent application goes through the examination process, an initial patent application may be divided into two or more related applications that share some of the claims of the original application. Using the consolidation described above, these patents will also be included in a single patent family. As a result, counts of environmental innovations using the OECD patent families will result in smaller counts than found in papers using data from single patent offices for three reasons:

1) Only innovations of high enough quality to justify pursuing protection at the EPO, USPTO, and JPO are included.

2) Several Japanese patents containing single claims will be consolidated into a single family.

3) Related patent applications at a single patent office are combined into a single family.

These first two reasons are clearly advantages of the TPF database, as they impose standards of quality and uniformity of patent laws across the countries in the database. The third reason, however, may offer one disadvantage. Applicants may occasionally file overly broad initial applications, hoping to see what claims will be upheld by the patent office. At times, the patent office may decide that the initial

---

[9] For example, the German patent office receives more applications from German inventors than American inventors each year, although the level of R&D is significantly higher in the U.S. than in Germany.

[10] Of course, we would still expect to see underrepresentation of inventors from countries outside of the Triadic countries.

application is too broad, and needs to be modified into two or more smaller patents (perhaps with additional claims added, known as a *continuation-in-part*). It would be valid to assign both patents resulting from such a continuation to the same priority date of the original application, as that date corresponds most closely to when the inventive activity took place. However, in cases where the patent office requested that the initial application be divided, it is also clear that the knowledge embodied in the initial application represented more than one patentable invention. In essence, the innovations represented by a single TPF are larger than the innovations represented by any single patent. While the standardization that results from using the TPF definition has its advantages, this is important to remember when comparing results drawn from this database to results from other studies.[11]

III. Working with Patent Data

To use the patent data included in the TPF (or elsewhere) to study environmental innovations, several basic steps are needed. First, one must decide on the patents to be included in the analysis. Typically, this begins with a decision to investigate a specific technology, and is followed with research necessary to identify patents pertaining to this technology. Once these patents have been identified, the patents must be organized for analysis. This involves sorting the patents by both date and source. I discuss each of these steps in greater detail below.

A. Choosing Relevant Technologies

The first step in any research project using environmental patent data is to decide on the technologies to be studied. Ideally, one wants technologies where there is a clear link to the policy in question. For example, for work on NO_X regulations for coal-fired power plants (Popp 2004), I identified specific technologies both for cleaning emissions, such as selective catalytic reduction (SCR), and for reducing the formation of NO_X in the combustion process, such as low NO_X burners. When searching for relevant patents, it is important to avoid searches that are too broad (by including too many irrelevant patents). For example, simply searching for all patents containing keywords such as "NO_X emission reduction" would also result in many patents pertaining to catalytic converters for automobiles, which are affected by a separate set of regulations. Fortunately, using the methods described below, these patents can be identified separately.

B. Choosing Relevant Patents

Once desired technologies have been chosen, the next research step is to identify relevant patents. This can be done by searching for patents in specific patent classifications, and/or searching for relevant key words. I discuss the specifics of each below, with particular attention paid to searches relevant to environmental technologies.

When patents are granted, they are given technology classifications and sub-classifications by various patent offices. These classifications can be used to identify all relevant patents for a given technology. The TPF database contains three potential classifications for each patent. First, the database

---

[11] An examination of the raw data suggests a related disadvantage. Some consolidated families capture a broad range of patents. For example, patent family #30 includes 5,038 patents. While one of these patents has the title "SYSTEM AND PROCESS FOR REMOVAL OF POLLUTANTS FROM A GAS STREAM" it also includes patents with such varied titles as "GOLF SWING TRAINING SYSTEM", "LEASE RENEWAL SERVICE", and "COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMORS". Moreover, while the environmental patent application was published in 2001, the date for the first patent in this family is 1927. Thus, using the first priority date would grossly misstate when the environmental innovation took place.

includes US Patent Classifications (USPC) placed on US patents by the USPTO. Second, the database contains International Patent Classifications (IPC) placed on the patent by both the EPO and the USPTO. Note that the IPC classifications assigned to patents in the same family may differ by patent office. The EPO uses the IPC as its main classification system. However, the USPTO relies on the USPC system. It assigns IPC classifications to patents based on a concordance between the USPC and IPC. However, these two systems differ in fundamental ways. The USPC uses a 'function-oriented' classification system, whereas the IPC uses an application-oriented approach. As such, most researchers use the IPC classification system when searching for patents.

In addition, since the mid-1990s, the EPO has made use of the European Classification System (ECLA). This classification scheme is based upon the IPC, but includes classifications at a greater level of detail. This additional detail allows for separate identification of technologies based on the pollutants they control. For example, IPC classification B01D 53/86 includes catalytic processes for pollution control. ECLA class B01D 53/86F2 specifies catalytic processes for reduction of NO_X, and B01D 53/86B4 specifies catalytic processes for reduction of SO_2. These classes can also help rule out irrelevant technologies. Using the example of catalytic converters from above, ECLA class B01D 53/94 and F01N 3/08B2 relate to catalytic converters for autos, and were thus not included in my study. Moreover, as new classifications are added, the EPO updates the ECLA of older patents in its database. As EPO patents in the late-1990s include these extra letters at the end of the IPC record, the TPF database appears to use the ECLA for these more recent patents. However, it does not appear to include ECLA assignments given to older patents after their initial publication.

Because the TPF database includes original patent classifications, care must be taken when choosing patent classifications to ensure that they are valid for the entire timeframe being searched. Fortunately, IPC manuals include a record of when each classification was added. The IPC is currently on its seventh edition. Revisions are typically made every five years. However, the seventh edition has been in effect since January 1, 2000, and its use has been extended until December 31, 2005. Version six was used from January 1, 1995, to December 31, 1999. Definitions of classifications that have been added since the first edition include numbers next to their definition stating when the class was added. For example, the entry for purification of waste gases appears as:

B01D 53/34   PERFORMING OPERATIONS; TRANSPORTING/ PHYSICAL OR CHEMICAL PROCESSES OR APPARATUS IN GENERAL/ SEPARATION/ Separation of gases or vapours; Recovering vapours of volatile solvents from gases; Chemical or biological purification of waste gases, e.g. engine exhaust gases, smoke, fumes, flue gases, aerosols/Chemical or biological purification of waste gases [3,6]

The bracketed numbers indicate that this listing was added and/or changed in version 3 (which took effect in 1980) and version 6 (which took effect in 1995). The modification that occurred in 1996 was the creation of additional subclasses that fall below B01D 53/34.[12] For example, it is now possible to distinguish between technologies that treat NO_X or SO_2:

B01D 53/50   PERFORMING OPERATIONS; TRANSPORTING/ PHYSICAL OR CHEMICAL PROCESSES OR APPARATUS IN GENERAL/

---

[12] Note that the IPC, as well as the USPC, use a hierarchal system. Hierarchies are represented by dots on the list of IPC definitions. B01D 53/34 has just a single dot. The classes appearing below it, such as B01D 53/50 (defined below) have multiple dots, signifying that they are a subset of B01D 53/34. In contrast, as B01D 53/32 "Controlling by gas-analysis apparatus by electrical effects other than those provided for in group B01D 61/00," also has a single dot, B01D 53/34 is not a subset of B01D 53/32.

	SEPARATION/ Separation of gases or vapours; Recovering vapours of volatile solvents from gases; Chemical or biological purification of waste gases, e.g. engine exhaust gases, smoke, fumes, flue gases, aerosols/Chemical or biological purification of waste gases/Removing components of defined structure/Sulfur compounds/Sulfur oxides [6]
B01D 53/56	PERFORMING OPERATIONS; TRANSPORTING/ PHYSICAL OR CHEMICAL PROCESSES OR APPARATUS IN GENERAL/ SEPARATION/ Separation of gases or vapours; Recovering vapours of volatile solvents from gases; Chemical or biological purification of waste gases, e.g. engine exhaust gases, smoke, fumes, flue gases, aerosols/Chemical or biological purification of waste gases/Removing components of defined structure/Nitrogen compounds/Nitrogen oxides [6]

However, note that these new classifications are only available on patents published since 1995. In addition, note that similar issues arise when using the USPC. However, the USPC does not list the dates of revisions. Thus, care must be taken to ensure that any US classifications chosen span the timeframe studied.

Unfortunately, the above example suggests that detailed searches of the TPF may not be possible for all but the most recent of years. However, the database offers one potential alternative. The database also includes English titles for patents published by the EPO and USPTO. Suppose, for example, that one wanted to identify all relevant NO_X control patents. While all patents in B01D 53/56 would be appropriate, this would only include patents published since the sixth IPC revision in 1995. Unfortunately, extending the search to include B01D 53/34 would include many patents related to pollution control, but not specifically to NO_X. However, combining this search with a search of titles can remove patents in this broader class that are not directly related to NO_X control. In Section IV I evaluate the potential for such combined search strategies.

Based on this information, how should one proceed to identify relevant classes? In my work, I have typically used the following strategy. I begin by obtaining information on appropriate technologies from various publications. For example, for my work on NO_X and SO_2 patents, I used documents on pollution control from the International Energy Agency's (IEA) Clean Coal Research Centre. For my work on energy patents, I identified relevant energy efficient technologies through documents published by the U.S. Department of Energy. Once relevant technologies have been identified, I use keyword searches to identify potentially relevant patents.[13] Individual documents were viewed to determine which were, in fact, patents pertaining to pollution control. I keep a tally of the classifications these patents make, so as to generate a list of classifications that occur frequently in the data. Then, I do two searches of these classifications: one to identify all patents in these classes, and a second in combination with the relevant keywords. Using this information, I identify classifications for which the majority of patents do pertain to the desired technology.

This last step can be particularly important. For example, my work with NO_X patents includes technologies used to modify the combustion process to reduce the formation of NO_X. However, many patents in potential combustion classifications pertain to other aspects of combustion (e.g. improving horsepower). As such patents should not be expected to respond to environmental regulation (or, even worse, may respond negatively), including such classes in my work would understate the links between environmental policy and innovation. Thus, given a choice between defining the search too broadly or

---

[13] In my previous work, these keyword searches include both the patent title and abstract. Unfortunately, the TPF only includes titles.

too narrowly, it is preferable to design a narrow search. Although the actual counts of patents produced may be low, the time series evidence linking innovation to environmental policy will be more accurate.

C. Organizing the Data – Choosing the Date

Once a suitable set of patent classes has been identified, how should the data be organized? Of concern here is both what date should be assigned to the patent, and to what country should it be attributed. Regarding dates, the appropriate date is the priority date of the patent. Much economic research (e.g. Griliches 1990) has found that patents sorted by their application date best correlate with R&D activities. Since the priority date is the earliest application date for the family of patents, this date is most closely associated to when the R&D took place.

One potential issue with the triadic family database is that consolidated patent families may include more than one priority date. In the entire database, just under one-half of the patent families include more than one priority year.[14] If each patent in the family were recorded separately, inventions would be counted at two separate priority dates. Assuming that continuing R&D efforts were made between filing the first patent application in the consolidated family and the patent application with the most recent priority date, patent counts based on the earliest priority date will under-represent the level of R&D that occurred in subsequent years. Given that most patent families have little or no gap, and that prior economic research finds close correlations between patent application dates and R&D activity, using the first priority date in the data makes sense. However, the issue of multiple priority years is likely to slightly dampen patent counts for more recent years.

A more important problem when using priority dates, either from the TPF or from other patent databases, is truncation bias. A patent family is not recorded until a patent document appears at the EPO, JPO, and USPTO. For the EPO and JPO, the necessary document is a patent application. These applications are typically published 18 months after the priority date. However, for patents that enter the EPO or JPO with a PCT application, there will be an additional delay. Applicants typically apply for protection in their home country first. Recall that inventors have one year from the priority date to apply for protection elsewhere. If the applicant chooses to file a PCT application through the World Intellectual Property Organization (WIPO), rather than filing directly in other nations, they have an additional 18 months before the PCT application must be transferred to various national offices for examination. During this time, a preliminary examination is done on the PCT application, thus providing the applicant with more information on the potential value of applications in various countries. In these cases, 30 months will pass between the initial priority date and the transfer of the application to the EPO and JPO. Assuming the TPF currently includes all patent applications published in 2004, this implies the database is only complete through the year 2001.

There is, however, an additional complication. Until 2000, patents were not published in the United States until they were granted. Thus, the US patent data in the TPF database comes from granted patents, rather than patent applications. Typically, it takes 2-3 years for a patent application to go through the examination process and be granted in the US (Popp, Juhl, and Johnson, 2004). However, this distribution is quite skewed. Between 1976 and 1997, five percent of patents took over five years to be processed.

Because US patents are not published until granted, we can never be sure that all patents with a given priority year have completed the examination process in the US. As such, research using US patent

---

[14] In the complete database, 53% of patent families have no gap between their earliest and latest priority year. 87% have a gap of two years or less, and 97% have a gap of five years or less. However, the tail is long, with gaps as long as 100 years observed.

data typically scales the data to account for grant lags in the US. Scaling typically proceeds as follows. First, one obtains a distribution of grant lags for patents that have been granted. If 50% of patents are granted within two years, this means that 50% of patents take longer. Thus, a count of patent applications from two years prior to the current date must be doubled to account for the 50% that have yet to be granted. In general, if x represents the cumulative percentage of patents granted by a lag of t years, $100/x$ represents the scaling factor that should be applied to patent counts from t years prior to today. This scaling factor typically falls to low levels (e.g. less than five percent) after approximately four years. Table 1 presents the grant lag for all patents granted by the USPTO between 1976 and 1997.

Unfortunately, while this adjustment is appropriate when studying US patents in isolation, it isn't clear that it applies in the case of the TPF database. Simply knowing the number of US patents is not enough. Rather, we need to know whether these patents also applied for protection abroad. Consider counts of US patents from 2000 and earlier, as these are the years with complete EPO and JPO data. While it is possible to scale patent counts from US data to get more complete counts for these years, we cannot ascertain whether the scaled patents also applied for protection in the EPO and JPO.[15] As such, the scaling techniques commonly used cannot be directly applied to the TPF database. Fortunately, since less than 10 percent of patents take longer than four years to be granted, and only 5 percent take longer than five years, any truncation bias will be small. Moreover, given that all US applicants who also file abroad now have their application published after 18 months, this problem will eventually go away in the TPF database.[16]

D. Organizing the Data – Choosing the Country

Regarding the country of origin of the invention, there are three possible choices. One is the priority country of the patent. This is the country where the initial patent application was filed. As such, it indicates the primary focus of the economic activity represented in the patent. Typically, the priority country is the home country of the inventor. However, as Dernis and Kahn (2004) note, this is not always the case. For example, many patents from Canadian inventors have a US, rather than Canadian, priority.

As such, the inventor or assignee countries are better indicators of the source of the patent. Each patent contains information on the inventor, including his or her country. In addition, most patents also record an assignee. The assignee is the person who has the legal rights to the patent. For example, if an employee of a firm receives a patent for work done on the job, the employee will be listed as the inventor, and the firm will be listed as the assignee. Thus, while inventor and assignee country will often coincide, there will be differences, particularly in the case of multinational countries. The assignee country will indicate who has economic control over the patent. However, the inventor country signals where the research was actually done. Moreover, not all patents list an assignee at the time of publication. For these reasons, using inventor countries to sort patents by country is preferred.

IV. Evaluation of the TPF Database

The preceding pages note both advantages and disadvantages to the TPF database. Before proceeding to propose research strategies for using the TPF database to evaluate environmental innovation, it is useful to first compare how patent counts constructed using the TPF database compare to counts using other sources, such as data from individual country patent offices. Other work comparing

---

[15] While the "nowcasting" included in the TPF database adjusts for this by looking at previous trends and accounting for growth in the probability of applying abroad, it is unlikely that such extrapolation would be reliable in a small subset of the data such as environmental patents.

[16] It will, however, still be a problem when using data solely from the USPTO, as applicants that do not file abroad can still choose to keep their application secret.

environmental patent counts across countries, such as Lanjouw and Mody (1996) and Popp (2004) use patent data from individual national patent offices. These data differ from the consolidated patent family database, as they include patents (most often from domestic inventors) that were filed at the national patent office, but not filed in other countries. Thus, patent counts from national patent offices will be larger than patent counts generated from the TPF database. Thus, one important question is whether the subset of patents filed at the EPO, USPTO, and JPO are an accurate gauge of inventive activity at the national level.

For this purpose, I include four figures comparing patent counts obtained with the TPF to patent counts found in Popp (2004). That paper uses patent data to compare innovation in the U.S., Japan, and Germany for NO_X and SO_2 pollution control technologies for coal-fired power plants. Of particular concern is whether the likelihood of patenting abroad varies across time, perhaps in response to domestic policy influences. Looking at patents pertaining to NO_X and SO_2 emissions control at coal-fired power plants, Popp (2004) finds that patenting trends respond to regulatory changes in the inventor's home country, but not to regulatory trends abroad. For example, NO_X patents from German inventor peak in the mid-1980s, in response to stringent NO_X standards passed at that time. Interestingly, this peak is not only observed for patents from German inventors granted in Germany, but also for patents granted to German inventors by the USPTO and JPO as well. However, there is no peak in patents assigned to German inventors, either at home or abroad, in response to changes in environmental policy in the US or Japan.

This suggests that patent family data should follow similar trends to national patent counts. Unfortunately, this paper also shows that the likelihood of German inventors of NO_X pollution control technologies filing for protection abroad also peaked in the mid-80s. The paper suggests that this occurs because the first wave of German NO_X control inventions were particularly innovative, as they responded to NO_X standards that were lower than elsewhere in the world. The same does not hold for SO_2, as the SO_2 standards passed in Germany were not significantly more restrictive than standards elsewhere. As a result, similar data collected from the TPF database would suggest a stronger response by German inventors to NO_X regulations than to SO_2. Moreover, as I discuss below, German inventors in the mid-80s were more likely to file for protection in Japan than in the US, since Japanese regulations were more similar to those in Germany.

Figures 1 and 2 compare counts of German and US patents pertaining to technologies designed to remove NO_X from the emissions stream. Each includes counts of patents assigned to German or US inventors, with patents sorted by their earliest priority date. Four separate time series are included:

Popp (2004): This includes patents for post-combustion treatment of NO_X emissions, taken from Popp (2004). The patents were identified using ECLA classifications from the EPO's website. This website includes re-classifications made to patents previously published. These counts are based on individual patent office data – that is, all US patents granted to American inventors, and all German patents granted to German inventors. For patents with multiple inventors, the country is based on the first inventor listed on the patent.

NO_X Only: This includes relevant patents taken from the TPF database. These are patents with either German or American inventors. For patents with multiple inventors, a fraction of the patent is assigned to each country, as described in Dernis and Kahn (2004). This search only includes IPC classifications directly related to NO_X control (B01D 53/56, 53/60, and 53/86[17]). However, these classes begin in 1995, so there are few patents in early years.

---

[17] B01D 53/86 refers to removal using a catalyst. It may include other pollutants, but is very important for NO_X removal, where the primary technique is Selective Catalytic Reduction.

Air Pollution Control: In addition to the patents in **NO$_X$ Only**, this also includes general air pollution patents (B01D 53/34).

Control + Title: In addition to the patents in **NO$_X$ Only**, this also includes general air pollution patents (B01D 53/34) that also have relevant keywords in its title.[18]

The most striking thing to note is that the counts from a single patent office (Popp 2004) are nearly always greater than the counts taken from the TPF database. The exception is for the time series including general US air pollution control patents, suggesting that such a search is too broad, and erroneously picks up other types of air pollution control patents. As noted earlier, this difference is expected, since the TPF uses an additional quality filter, and only includes patents filed at the EPO, USPTO, and JPO. However, as a result of this additional filter, the TPF data misses some of the time-series variation in the data. For example, in Popp (2004) German NO$_X$ patents peak in 1985, one year after passage of the first NO$_X$ regulations for coal-fired power plants in Germany. However, none of the German counts, including the count for general air pollution control, have such a peak.[19]

One reason for this may be that few German patents file in all three countries.[20] Of the German patents from 1985 included in Popp (2004), 46 percent have a Japanese family member. However, only 17 percent have a US family member. While Japan also had stringent NO$_X$ regulations at this time, the US did not. As such, German inventors appear to not value protection in the US for these inventions. In contrast, when US NO$_X$ patents peak in 1990, 71 percent also have German family members, and 69 percent have Japanese family members. As the US was the last of these three countries to pass stringent NO$_X$ regulations for coal-fired power plants, it makes sense that American inventors would also perceive Germany and Japan as a potential market for their innovations. Unfortunately, this suggests that patent counts using the TPF database are not only influenced by regulations in the inventor country, but also by regulations throughout the member countries of the database. As such, the TPF database is not appropriate for identifying research done by first movers in environmental regulation.

Similar problems are observed in Figures 3 and 4, which present the same time trends pertaining to patents for SO$_2$ control. Once again, patent counts from domestic patent offices are significantly higher. While a small peak in US SO$_2$ patents can be seen using the **Control + title** classification, the peak is not nearly pronounced as in the data from Popp (2004). Moreover, the peak in German SO$_2$ patents observed in 1984 is not found using any of the classification schemes with the TPF data.[21]

---

[18] The relevant keywords are: ' NOX ' OR ' NOx ' OR 'nitrogen dioxide' OR 'Nitrogen Dioxide' OR 'NITROGEN DIOXIDE' OR 'Nitrogen Oxide' OR 'nitrogen oxide' OR Title contains 'NITROGEN OXIDE'. Searches were done using SAS, which is case sensitive. Note that I include a space before and after the abbreviation NO$_X$ to avoid incorrectly identifying words containing the letters "nox", such as "noxious".

[19] While this may be expected for the count using only the recent NO$_X$ classifications, since they do not enter the IPC until 1995, it is surprising for searches that include more general air pollution control technologies, either with or without the additional key word search.

[20] Popp (2004) includes 857 German patents assigned to German inventors. Of those, only 118 have family members in the EPO, US, and Japan. Note that the definition of patent family used here comes from the INPADOC patent database, and is less restrictive than the consolidated patent family definition.

[21] As verification of the data in Popp (2004), note that a recent working paper by de Vries and Withagen (2005) finds similar trends and magnitudes for sulfur dioxide abatement patents. Their search strategy begins with keyword searches of patent abstracts, which identifies over 5,000 relevant patents for 14 separate countries. They then examine the abstracts individually to verify the patents are appropriate. Their searches are broader, as they include technologies such as coal desulfurization, where as Popp (2004)

The above results suggest that if one aims to use patent counts as a measure of the *level* of innovative activity in a country, counts from national patent offices may more accurately represent the level of activity.[22] The TPF database potentially undercounts *national* innovative activity for two reasons.   First, patents only enter the TPF database if an application is filed at the EPO, USPTO, and JPO.   While this has the advantage of ensuring the patents are of high quality, the above example suggests that the lack of a market for a new environmental innovation in one of these countries (or in a set of European countries) will result in an undercount of innovative activity.   For the TPF counts to be an accurate time series measure of changes in innovative activity within a country, it must be the case that the likelihood of filing an application abroad is constant over time.   If not, changes in time series trends may result from changes in patenting incentives abroad, rather than from changes in response to environmental policy.   Because there are fewer foreign markets for technologies developed by countries that choose to regulate first, *the innovative levels of countries that regulate early, such as Germany for NO_X, are likely to be understated.*

Second, the TPF database only includes the original patent classifications.   Because of this, note that the counts of NO_X innovations were low during the 1980s, as the IPC classification specific to NO_X began in 1995.   Searching more broadly (e.g. all air pollution control patents) introduces additional error, as it includes in the search patents that are not specific to the technology.   In the above case, for example, using the broader air pollution class as a substitute results in the same patents being counted for both SO_2 and NO_X control.   As such, a preferable strategy would be to combine these broader classes with a keyword search.   Unfortunately, in the above examples, these keyword searches add only a few patents to the counts using only the SO_2- or NO_X-specific technology classes.[23]

V. Suggestions for Empirical Work

These findings suggest a couple of guidelines for selecting appropriate environmental policies to study using the TPF database.   First, the database is less likely to be effective for identifying patents that can only be identified using classifications recently added to the IPC.   This suggests that studying broader air or water pollution technologies is preferable to studying technologies that focus on specific industrial applications.   Second, and perhaps more troubling, is the concern that the patenting activity for first movers may be underrepresented.   While this suggests choosing policies that have widespread adoption,

---

only uses patents pertaining to scrubbers.   As such, the trends in patents are similar to Popp (2004), but the magnitudes are even higher.

[22] This assumes, for example, that equal effort is put into similar SO_2 and NO_X patents.   If more research effort is necessary to obtain the higher quality patents represented by the NO_X patents that also filed abroad than the SO_2 patents that were only of interest to the local market, patent family data would provide a more accurate guide to the level of innovation. While there is a rich literature on the correlations between patent counts and R&D spending n general, only one paper, Lanjouw and Schnakerman (2004) compares correlations of weighted versus unweighted patent counts to the level of R&D activity.   Their paper weights patents by citations and by the number of claims.   If firms actively pursue an R&D strategy that focuses R&D efforts towards the most productive results, using weighted patent counts should improve the correlation between patent counts and R&D.   In contrast, if differences in the quality of patents result from unanticipated random shocks to the R&D process, weighted patent counts will not improve the correlation.   They find the later to be true, suggesting that firms do not know *a priori* which research projects will be most productive.   However, care must be take extending the result of their work to the current scenario.   Differences in the patentability of the German NO_X versus SO_2 innovations result not just from random shocks affecting the quality of the research, but also from the need of German inventors to develop technologies allowing German plants to comply with NO_X regulations surpassing those in force elsewhere in the world.

[23] One reason for this may be that the database allows searches for titles only.   Patents with general titles (e.g. air pollution control device) that specifically mention SO_2 or NO_X in the abstract would not be identified.

such a strategy limits the potential to identify the effects of policy on innovation, since there will be little cross-sectional variation in the data.

These shortcomings suggest that the TPF database is not, by itself, an appropriate data source for evaluating the effect of environmental regulations on innovation. Patent data taken from national patent offices will be more appropriate. Counts from individual patent offices are needed to ensure that the patenting activities of early regulators are properly represented. The appendix to this report discusses three potential sources of data that could be used. In such a framework, the added information about patent quality from the TPF could be used as an important control. Because patents in the TPF database have sought protection at three major patent offices, these patents can be viewed as the global "frontier" of knowledge for any technologies studied.

Thus, based on the example above, while an interpretation that *the level of inventive activity of German inventors* on NO_X control versus SO_2 control may be skewed by looking at patent family data, an interpretation on the value of the *contribution of German inventors to the global knowledge* stocks of NO_X and SO_2 emission control technologies is more accurately obtained using the TPF data. That is, one could argue that the fact that German NO_X inventions were not patented in all three TPF countries means that the inventions were not targeted at foreign markets.

This alternative interpretation of the TPF data can be particularly useful for controlling for supply-side influences on innovation. Using US patent data to estimate the effects of energy prices on energy-saving R&D, Popp (2002) shows that controlling for the current state of technology is important, as research opportunities *within a given technological field* experience diminishing returns over time. Similarly, in Popp (forthcoming), I discuss how not including controls for the current state of technology leads to lower estimates of the response of patents to environmental regulation in Brunnermeier and Cohen (2003) and Jaffe and Palmer (1997).

Pertaining to work on international comparisons of environmental policy and innovation, diminishing returns suggests larger responses to policy in countries that adopt regulation early, and potentially smaller responses to policy in countries adopting regulation later.[24] Thus, regressions examining the link between environmental policy and patents might take the following form:

$$Pat_{i,t} = f(Policy_{i,t}, KnowledgeStock_{t-1}, X_{i,t})$$

where i and t index country and time, respectively, *Pat* is an indicator of relevant environmental patents in country i at time t, *Policy* is a measure of environmental policy stringency, $KnowledgeStock_{t-1}$ represents the lagged stock of knowledge available for the technology in question, and X controls for other factors

---

[24] Whether late adopters do less innovation depends on whether international technology transfer takes place. This depends not only on trade between regulating countries, but also on whether technologies adopted in one country are suitable for adoption in a second country. For example, Popp (2004) shows that despite adopting stringent NO_X regulations later than Japan or Germany, US inventors still patent in response to stringent NO_X regulations passed in the US in the 1990s. Using patent citations, the paper presents evidence that these patents build on the knowledge embodied in earlier Japanese and German patents. Whether these results hold for other technologies would be an interesting research question.

influencing patenting trends in the country.[25] Note that, using the TPF, this stock need not be indexed by country, as the TPF can be considered a measure of *global* knowledge.[26]

Few papers include controls for such supply side measures, as obtaining such data can be difficult. For example, while I use patents weighted by the likelihood of citation in Popp (2002), I have not been able to apply the techniques used in that paper in other research, as the data requirements are too burdensome for a study of very narrowly defined technologies. While a time trend can be used in place of such stocks, a time trend does not capture variation in the stocks that occur due to policy shocks. For example, a time trend cannot identify the effect that first mover countries have on the global knowledge frontier. Using the TPF to measure the global technology base thus offers a useful alternative that could potentially be applied to a wide variety of circumstances.

For the above regression, it is important to consider how the dependent variable is measured. Since domestic inventors are more likely to be affected by environmental policy than foreign inventors, separate counts of patents for foreign and domestic inventors will generally be used. Patent counts will vary by country, due to differences in national patent laws and variation in country size. US counts will be high simply due to the size of the economy, and Japanese counts will be high due to the nature of Japanese patent law. If raw patent counts are used as the dependent variable, country fixed effects should be included to control for differences in the level of patenting activity across countries. However, unless the environmental policies included in the regressions vary across time as well as country, their effects cannot be separately identified using fixed effects.

While variables such as GDP and the number of scientists and engineers in a country control for economic factors influencing innovation levels across countries, these variables are not a perfect substitute for country fixed effects, as they do not control for differences in national patent systems. As such, a better solution is to define the dependent variable as the percentage of all domestic patent applications in country i from a given year, t, that are environmental patents. These percentages control for differences in the level of patenting activity across countries. Moreover, if changes in patent law lead to increases or decreases in the general level of patenting within a country, percentages will control for this growth, as both the numerator and denominator will increase.[27]

It is also useful to consider how to include environmental regulations in the above equations. Ideally, one would like specific measures of stringency, so that information on differences in the levels across countries can be used.[28] However, as Figures 1-4 show, the reaction to new command and control (CAC) regulation tends to be strongest when the regulation is first introduced. This is consistent with economic theory, which notes that CAC regulations provide less incentive for innovation than market-

---

[25] Examples of other controls that could be included are measures of human capital (which are unlikely to vary much in the countries studied using the TPF), government R&D efforts, and the percentage of GDP contributed to the industry in question (to control for the importance of the industry within each country).

[26] Indeed, for the US, Japan, and Europe, it cannot be indexed by country, since the database by definition contains patent applications filed at all three patent offices.

[27] For example, the rate of annual growth in patenting grew dramatically in the U.S. during the 1990s.

[28] If some countries have no regulation, these levels should be expressed as a percentage. For example, consider three countries. Country A has no limit on NO_X emissions. Country B has a limit of 50 ppm, and Country C has a limit of 100 ppm. B's regulation is more stringent than C's, but both are more stringent than A, which has no limit. Simply using the numerical limits, however, suggests $0 < 50 < 100$, which makes A look more stringent. Instead, define *relative stringency* as (minimum regulation)/regulation if a regulation exists, and 0 otherwise. This measure equals 0 for A, 1 for B, and 0.5 for B, which accurately captures the relative stringency of each country's regulation.

based policies. This is because CAC policies provide no reward for *exceeding* the standard, whereas firms that exceed the standard under a market-based policy are rewarded with lower emission fees or the opportunity to sell excess pollution permits. This suggests that a couple of alternatives may provide better fits when CAC regulations are enacted:

- Dummy variables for the years in which a policy is enacted. Of course, specification checks on the timing (e.g. a dummy for the year a policy is enacted, versus one for the year after) will need to be performed.

- Including *changes* in stringency levels, as opposed to the levels themselves. Again, lagged values can be included to check the timing of the response.

In addition, given the hypothesized differences between innovation incentives for CAC and market-based policies, comparing innovation levels for cases in which some countries make use of each policy type would be fruitful. Based on the above discussion, theory would predict that the innovative response to market-based policies should last longer. A challenge will be to include indicators for both market-based and CAC policies in the same regression. Of course, country and year specific dummy variables are a possibility. Another option would be to find the *equivalent* standard for a market-based policy. For example, if a market-based policy is put in place with a specific quantity goal (e.g. a specific amount of tradable permits), that quantity could be treated the same as a CAC limit. Then, an additional interaction term, including this level along with a dummy variable signifying a market-based policy is in place, could be added to the regression. Theory predicts that the coefficient on this variable would be positive – given two countries with similarly stringent policies, we'd expect more innovation in the country using a market-based standard.

Suggestions for Analysis

The empirical strategy discussed above suggests several characteristics desirable for any data chosen for analysis. I first list these characteristics below, and then suggest specific examples for potential research. In these examples, I highlight additional data (typically regarding environmental regulations) that would be needed for analysis. In particular, I focus on cases where OECD access to appropriate government agencies may be helpful to address specific data needs.

To identify the effect of regulations on innovation, the regulations must vary across time and/or across country. Regression analysis is most likely to be successful with policies that vary in both dimensions. Environmental problems for which different mechanisms are used in different countries would be of particular interest. For example, while most countries have used command and control regulation for NO_X, France and Sweden have both instituted a system of emission charges. Similarly, the Netherlands use effluent charges for water emissions. Kemp (1997) has studied the effects of these laws on technological change and diffusion. Extending that work in a comparative, cross-country setting is a natural extension.

Clearly identifiable technologies should be used to help firms (and/or consumers) comply with these regulations. For example, in response to the phaseout of leaded gasoline, new automobiles required catalytic converters. Moreover, refineries turned to technologies such as isomerization to remove lead from gasoline (Kerr and Newell 2003). Note that end-of-the-pipe abatement technologies involve product innovations, which are more likely to be patented than process innovations. Moreover, once appropriate technologies are identified, checks will need to be made to establish links between these technologies and patent classifications (or a combination of classifications and keywords) that are reliable *across the time period studied* (as described in section II). Schmoch (no date) provides some guidelines for patent searches pertaining to environmental technologies. However, the searches proposed are very broad, and

may be difficult to link to specific regulations. Simply developing a database of more narrowly defined environmental technologies should be seen as a valuable research contribution.

i. Selection of appropriate countries (and regulations)

As noted, variation both across time and across time is important to identify the effects of regulation on innovation across multiple countries. Given that the leading economies are likely to be early adopters of regulations, this suggests focusing research on a broad range of OECD countries, not just on the leading economies. As a specific example, research on the flows, if any, of environmental technology between the United States, Canada, and Mexico could be of great interest. One concern raised by opponents of the North American Free Trade Agreement (NAFTA) was that dirty industries could simply move to Mexico, where environmental regulations are weaker. At the same time, however, freer trade could improve the availability of environmental technologies in Mexico.

Patent data from the three countries could be used to ask whether freer trade does increase the flow of environmental technology to Mexico. Such research should examine both patenting trends in the US, Canada, and Mexico (e.g. do environmental patents increase in Mexico after passage of NAFTA, and are the US and Canada a major source of these patents?) and trends in environmental regulation in Mexico. For this, information on whether Mexican environmental regulations changed after enactment of NAFTA would be needed. This information should focus both on regulatory levels *and* on the amount of money spent by the Mexican government on enforcement of environmental regulations. This research would help assess both whether freer trade improves access to environmental technologies for developing countries, and whether such trade agreements give advantages to first movers in environmental regulation, such as the United States.[29]

ii. Research on different policy mechanisms

As noted, one useful research task would be to evaluate how different types of environmental policy affect innovation. Here, two pieces of information could be of particular use. First, to compare policies using different mechanisms (e.g. command and control versus market-based policies), a common metric is needed. Presumably, converting market-based mechanisms to a quantity measure will be more straightforward than converting command and control quantities to a market-based measure. For example, information on either *actual reductions* resulting from environmental taxes or on *target levels of reductions* that were used to justify market-based policy could be used to express the market-based regulation using quantity measures. One could then compare the effect of market-based measures to command and control measures by using an interaction term that includes both the target quantity and a dummy variable equal to one if a market-based measure was used.

Second, many regulations depend on a specific technology. For example, many environmental regulations in the United States require that the Best Available Control Technology (BACT) be used. Such regulation should direct research toward a specific technology. This could potentially lead to lock-in by reducing innovation incentives for other technologies. Taylor *et al.* (2003) examine patenting activity after the 1977 US Clean Air Act mandated use of scrubbers with 90% removal efficiency at US power plants. Not only did patents for SO_2 scrubbers increase after this mandate, but patents for alternative methods for reducing SO_2 emissions, such as clean coal technologies, decreased. As such, information from member countries on *specific* technologies mandated through environmental regulations would be useful. For example, if different countries mandate different technologies, one could ask whether such mandates slow the diffusion of other new environmental technologies.

---

[29] This work would complement the on-going project on globalization and 'green innovation' being undertaken under the Working Party on Global and Structural Policies. (See ENV/EPOC(2004)1).

In addition, even if regulations do not mandate a specific technology, regulators may consider the availability of key technologies before setting a regulation. Again, knowing whether such information has been used to set regulations could help focus research on both technologies that could be fostered and hindered as the result of environmental regulation.

Finally, environmental regulations are not the only policy mechanism used to induce innovation on environmental technologies. Government subsidies, either through direct research funding or using tax credits, are also important. Data on such policies, in addition to data on environmental regulations, would be very valuable. Of course, for econometric work, historical time series would be needed for multiple countries.

iii. Choice of appropriate technologies

Finally, appropriate technologies must be chosen. Of course, answers to the questions raised above help define what technologies could be included for analysis. Nonetheless, there are some broad categories of technologies and industries that, because of their nature, are particularly suited to analysis using patent data. The distinction between changes-in-production processes and more readily identifiable (but perhaps less innovative) end-of-pipe technologies is one important distinction.

One potentially fruitful area is the chemical industry. Compared to other industries, the chemical industry makes greater use of patent protection (Levin *et al*. 1987). This is because new chemicals are straightforward to patent (it is clear whether a chemical formula is new or copies an existing formula), and because chemicals could be easily copied without patent protection. Thus, it may be possible to identify patent classes pertaining to either new chemicals that are designed to replace chemicals being phased out (e.g. after the Montreal Protocol), as well as to identify patent classes for chemicals being regulated. Moreover, if firms change the chemicals used in the face of regulations (e.g. regulations on water effluent), identifying the chemicals involved may offer a window into process innovation, as opposed to the many end-of-pipe technologies that have been identified in other work.

V. Conclusions

The above report discusses the advantages and disadvantages of using patent data to evaluate environmental innovation, with a particular focus on the Triadic Patent Family database developed at the OECD. I identify two potential weaknesses of the database: it includes only original patent classifications, so that less detail is available for early years of the database, and patent families may be sensitive to the weakest environmental regulation among the Triadic Patent Family countries. Any research design using these data must be cognizant of these issues. However, because the TPF database can serve as a measure of the global knowledge frontier, using the TPF in combination with data from national patent offices would be an improvement over the existing literature. As such, I conclude with several suggestions for research that make use of country-specific patent data. Finally, it is important to note that, while I suggest a regression framework above, my previous experience with similar data suggests that finding a good regression fit will be difficult. Unfortunately, environmental regulations do not vary much within countries. Thus, much of the identification will be across countries. It may be that the number of countries to be included, which is more than I have used in previous work, will provide enough cross-sectional variation to overcome the lack of time-series variation. However, if this is not the case, more descriptive work, as in Popp (2004) is recommended.

REFERENCES

BRUNNERMEIER, SMITA B. AND MARK A. COHEN (2003), "Determinants of environmental innovation in US manufacturing industries," *Journal of Environmental Economics and Management*, 45(2), 278-293.

DE VRIES, FRANS P. AND CEES WITHAGEN (2005), "Innovation and Environmental Stringency: The Case of Sulfur Dioxide Abatement," CentER Discussion Paper #2005-18, Tilburg University.

DERNIS, HÉLÈNE AND MOSAHID KAHN (2004), "Triadic Patent Families Methodology," *STI Working Paper 2004/2*, Organisation for Economic Co-operation and Development.

GRILICHES, ZVI (1990), "Patent Statistics As Economic Indicators: A Survey," *Journal of Economic Literature*, 28(4), 1661-707.

JAFFE, ADAM B. AND KAREN PALMER (1997), "Environmental Regulation and Innovation: A Panel Data Study," Review of Economics and Statistics, 79(4), 610-19.

KEMP, RENÉ (1997), *Environmental Policy and Technological Change: A Comparison of the Technological Impact of Policy Instruments*, Edward Elgar Publishing, Cheltenham, UK.

KERR, SUZI AND RICHARD G. NEWELL, (2003), "Policy-Induced Technology Adoption: Evidence from the U.S. Lead Phasedown," *Journal of Industrial Economics* 51(3),317-343.

LANJOUW, JEAN O. AND ASHOKA MODY (1996), "Innovation and the International Diffusion of Environmentally Responsive Technology," *Research Policy*, 25, 549-571.

LANJOUW, JEAN O., ARIEL PAKES AND JONATHAN PUTNAM (1998), "How to Count Patents and Value Intellectual Property: Uses of Patent Renewal and Application Data," *The Journal of Industrial Economics*, 46(4), 405-433.

LANJOUW, JEAN O. AND MARK SHANKERMAN (2004), "The Quality of Ideas: Measuring Innovation with Multiple Indicators," *Economic Journal*, 114(495), 441-465.

LEVIN, RICHARD C., ALVIN K. KLEVORICK, RICHARD R. NELSON, AND SIDNEY G. WINTER (1987), "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3, 783-820.

ORDOVER, JANUSZ A. (1991), "A patent system for both diffusion and exclusion," *Journal of Economic. Perspectives*, 5 (1), 43-60.

POPP, DAVID (forthcoming), "Lessons from Patents: Using patents to measure technological change in environmental models," forthcoming in *Ecological Economics*.

POPP, DAVID (2004), "International Innovation and Diffusion of Air Pollution Control Technologies: The Effects of NO_X and SO_2 Regulation in the U.S., Japan, and Germany," *NBER Working Paper #10643*.

POPP, DAVID (2003), "Pollution Control Innovations and the Clean Air Act of 1990," *Journal of Policy Analysis and Management*, 22(4), 641-660.

POPP, DAVID, (2002), "Induced innovation and energy prices," *American Economic Review*, 92, 160-180.

POPP, DAVID, TED JUHL, AND DANIEL K.N. JOHNSON (2004), "Time in Purgatory: Determinants of the Grant Lag for U.S. Patent Applications," *Topics in Economic Analysis & Policy*, 4(1), Article 29 http://www.bepress.com/bejeap/topics/vol4/iss1/art29.

SCHANKERMAN, MARK. AND ARIEL PAKES (1986), "Estimates of the value of patent rights in European countries during the post-1950 period," *Economic Journal*, 96, 1052–1076.

SCHMOCH, ULRICH (no date), "Definition of Patent Search Strategies for Selected Technological Areas: Report to the OECD."

TAYLOR, MARGARET R., RUBIN, EDWARD S., AND DAVID H. HOUNSHELL (2003), "Effect of government actions on technological innovation for SO_2 Control," *Environmental Science & Technology*, 37(20), 4527-4534.

Table 1 – Grant Lag Distribution for U.S. Patents

grant lag (in years)	Number	Cumulative Frequency	Cumulative Percentage	Scaling factor
1	112007	112007	6.77	14.771
2	777337	889344	53.77	1.860
3	444086	1333430	80.63	1.240
4	160367	1493797	90.32	1.107
5	73632	1567429	94.77	1.055
6	37466	1604895	97.04	1.031
7	19764	1624659	98.23	1.018
8	11276	1635935	98.92	1.011
9	6137	1642072	99.29	1.007
10	3689	1645761	99.51	1.005

The table shows the lag, in years, between the first related U.S. patent application and the publication of a granted patent, at the USPTO. The data includes all patents issued between 1976 and 1997.
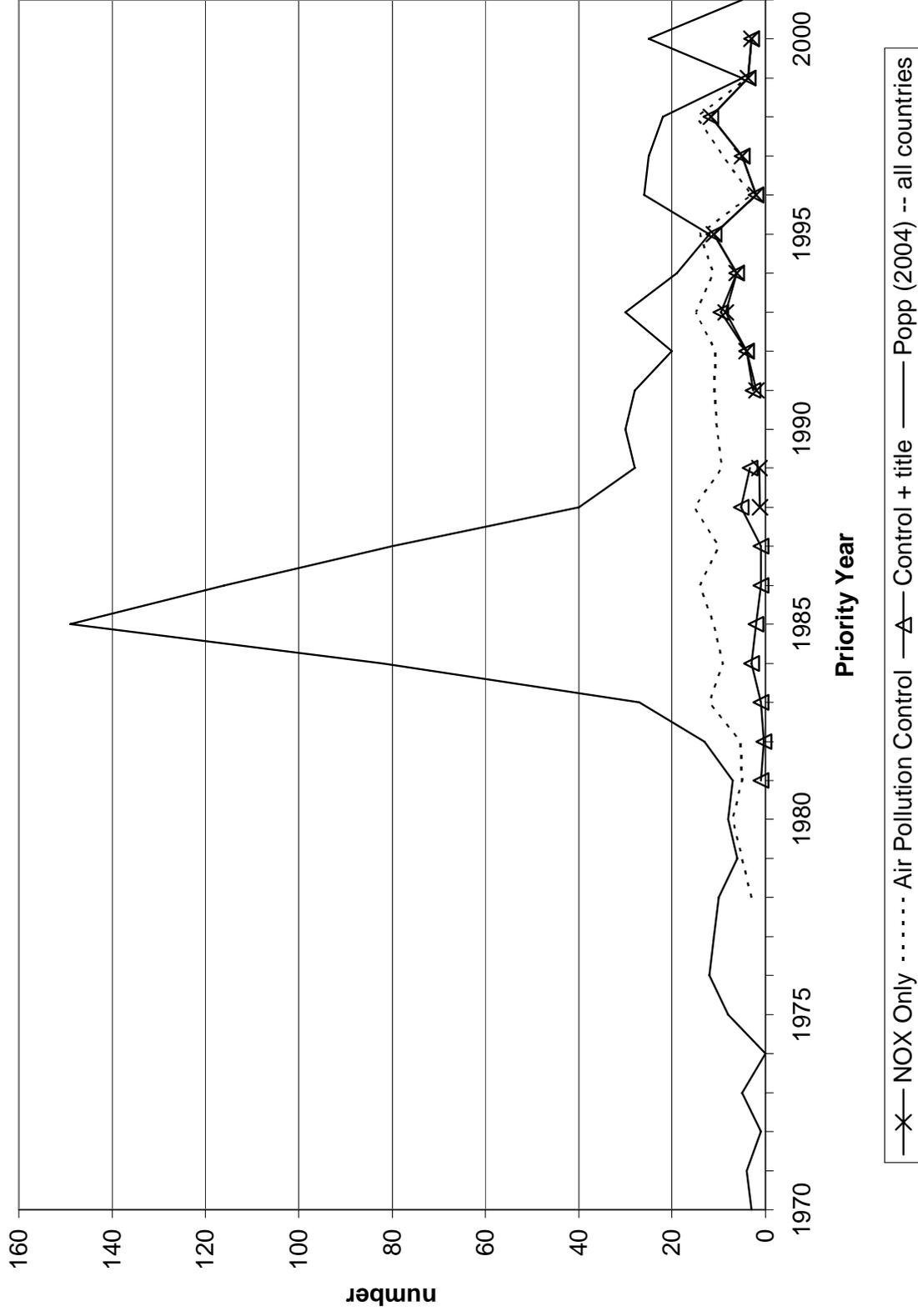
Figure 1 -- German NOX Post-combustion Patents

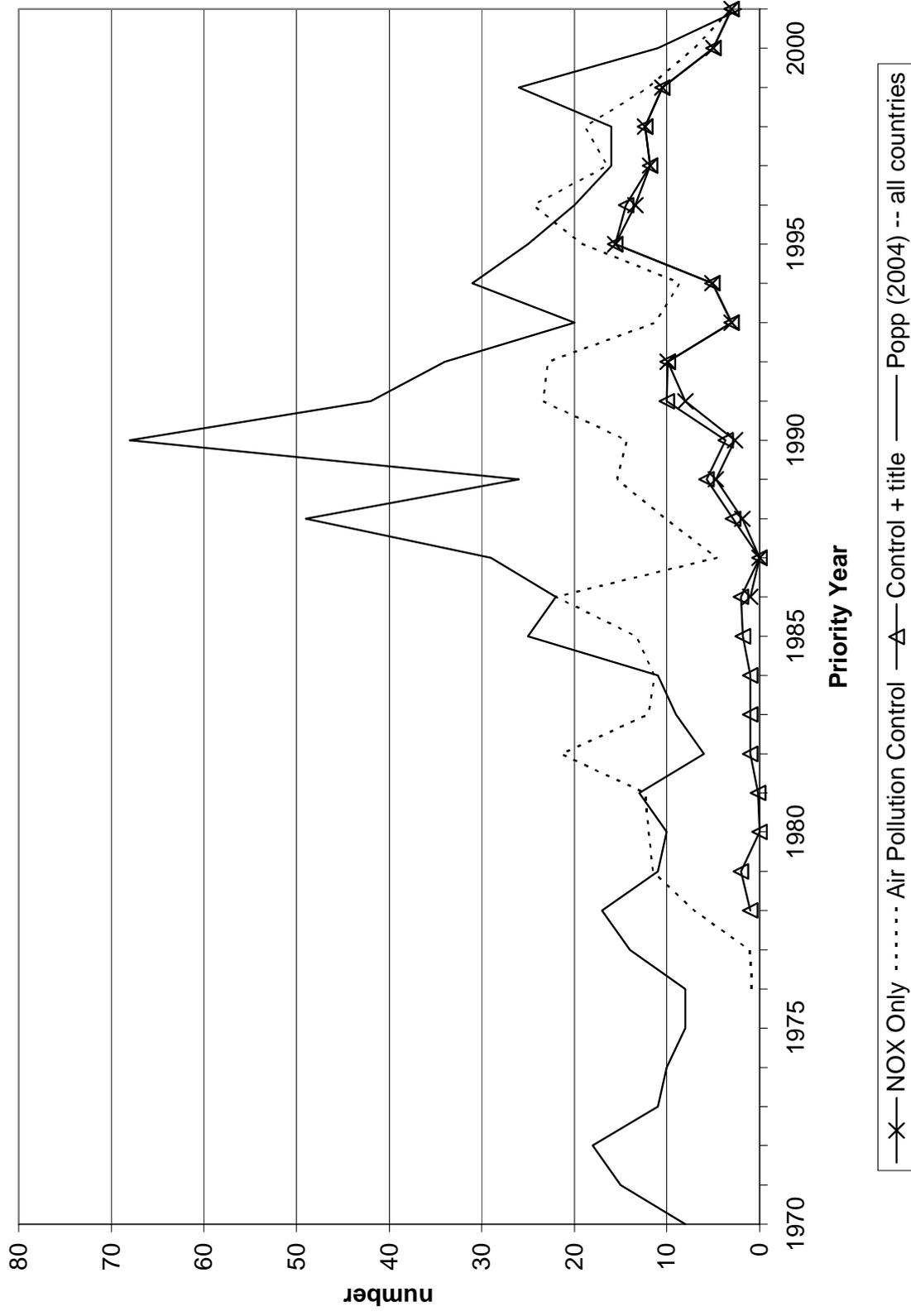— NOX Only · · · · · Air Pollution Control —△— Control + title —— Popp (2004) -- all countries

23

Figure 2 -- US NOX Post-combustion Patents

24

Figure 3 -- German SO2 Patents

Priority Year

number

SO2 Only ⋯⋯ Air Pollution Control —△— Control + title —— Popp (2004) -- all countries
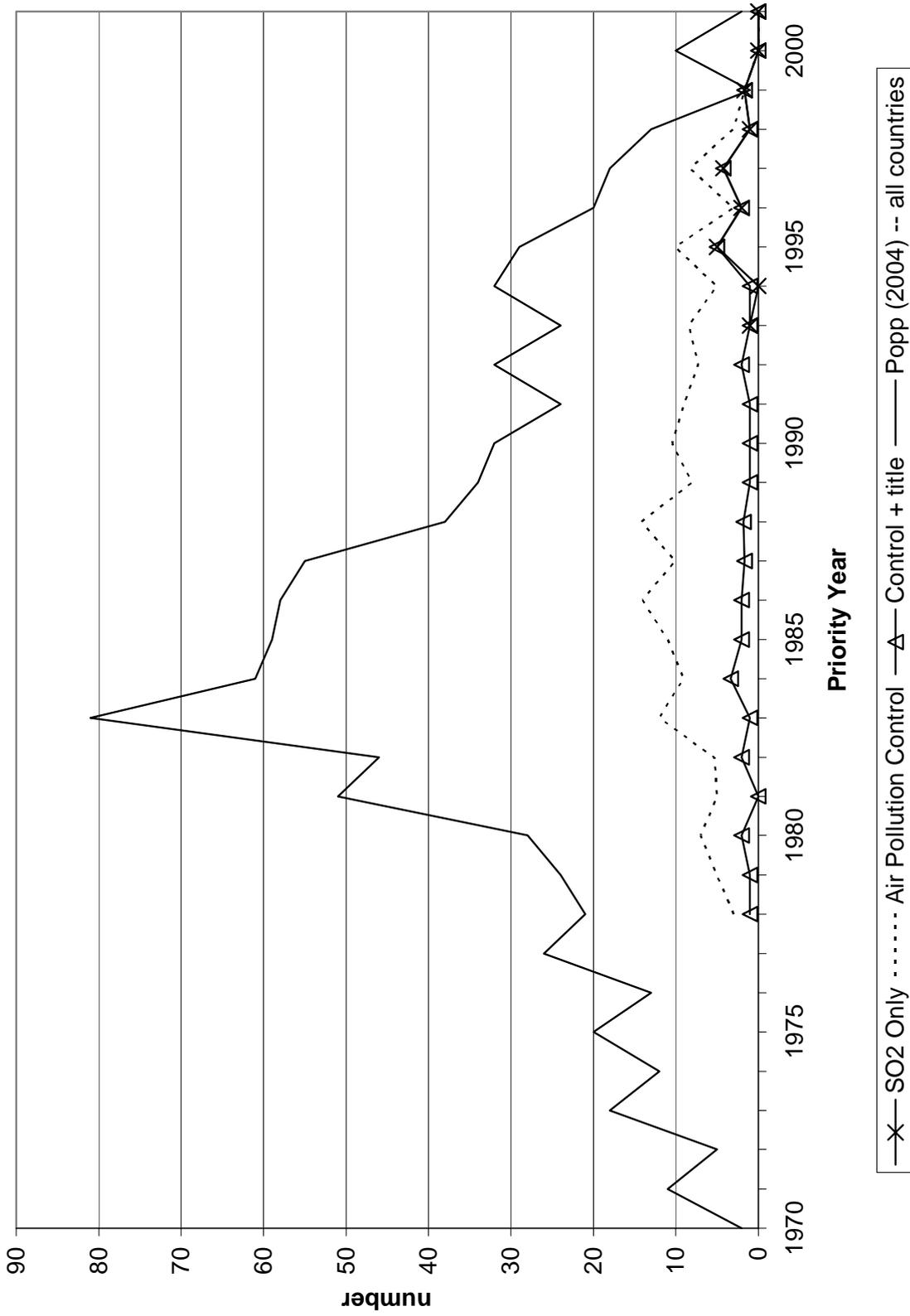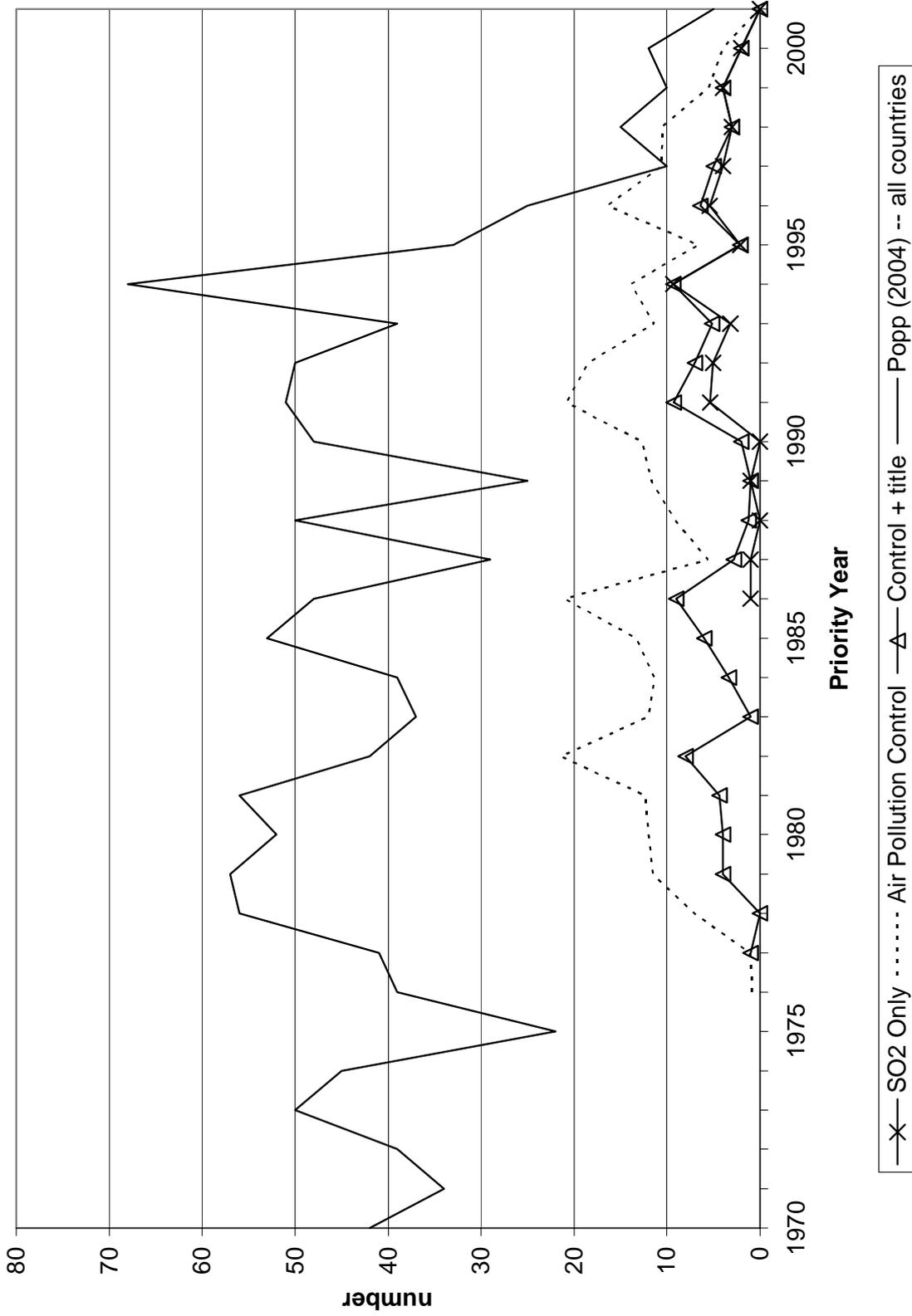
25

Figure 4 -- US SO₂ Patents

26

APPENDIX – SOURCES OF PATENT DATA

As noted in this report, the data in the Triadic Patent Family database are not sufficient to analyze trends in environmental innovation. Suitable data sets should provide coverage at the country level, and should allow searches over a broad range of categories and keywords. Below I discuss three data sets, two publicly-available and one commercial, that meet these criteria.

The first data set is esp@cenet. It is a searchable on-line database provided by the European Patent Office (EPO) that can be accessed at http://ep.espacenet.com/. While the website includes access to four patent databases (EP, WIPO, Japan, and worldwide), it is the worldwide database that is of most use for this research.[30] The worldwide database currently includes information on patents from 78 patent offices.[31] The range of data availability varies by country, so researchers should check the availability before beginning research to determine the time frame suitable for analysis. The most relevant data limitations are that only recent data is available for some developing countries, and European Patent Classifications (ECLA) are not readily available for many non-EPO countries (with the US being an exception). As noted in the report, this classification scheme is based upon the often-used International Patent Classification system (IPC), but includes classifications at a greater level of detail. More importantly, as new classifications are added, the EPO updates the ECLA of older patents in its database.

Researchers can use the esp@cenet search pages to generate a list of patents meeting a specific criterion (e.g. all patents from a specific country and classification in a given year). These searches can be done using the advanced search page. To specify a patent country, simply enter the two letter country abbreviation in the publication number field. This will yield all patents published in a specific country. Unfortunately, it is not possible to limit searches to all patents with inventors from a specific country. The closest approximation is to use the priority number. Since most applicants first file in their home country, entering the same country in both the publication number and priority number will approximate the number of domestic patents. However, this will not yield a perfect count, as some patents have multiple priority numbers. For example, a search of US publications published in 2000 in ECLA class B01D 53/56 yields 35 patents. Limiting this to US priority numbers results in 14 patents. These 14, however, include one patent from Japanese inventors with an earlier Japanese priority number, as well as a patent from Canadian inventors with a US priority number.

One other limitation to esp@cenet searches arises if the technology being searched includes several possible ECLA codes. Only four search terms can be combined at a given time. If more than 4 classifications are being searched, multiple searches will need to be run. However, these searches may yield duplicate results. For example, searches for ECLA B01D 53/34 and for ECLA B01D 53/56 may *each* yield a patent that includes both ECLA. Thus, simply adding the search results will lead to double counting. This can be avoided by saving the individual patent numbers generated from the search, and then manually removing duplicates. Unfortunately, one drawback of the esp@cenet site is that data on

---

[30] Note that the worldwide database includes data on patens from the EPO, WIPO, and Japan. One difference is that the worldwide database does not include translations of patent abstracts from Japan, whereas the Japanese database does.

[31] An updated list of available data can be found at http://ep.espacenet.com/ep/en/helpV3/detailedcoverage.html.

individual patents cannot be downloaded. Thus, this would be done by first manually cutting and pasting the search results into a separate program, such as Microsoft Excel.

An alternative source that does allow users to download descriptive data is the Delphion on-line database.[32] This database, available only by subscription, allows the user to download multiple patent records in machine-readable formats. Descriptive data available include the priority, application and issue dates for each patent, the home country of the inventor, and data on patent families. Unfortunately, updated ECLA classifications are not available from Delphion. In Popp (2004), I matched this data to lists of patents from relevant ECLA by first getting lists of relevant patent numbers from esp@cenet. I then downloaded patents from Delphion that fell in related IPC classifications, and merged these data with the lists of patent numbers to restrict the analysis to those falling in the relevant ECLA classes. Various counts can be obtained from the raw data using standard data management software, such as SAS.

Finally, like the esp@cenet data, the Delphion database does not include detailed descriptive data for Japanese patents. Some additional data on these patents can be obtained from the Japanese Patent Office (JPO) website.[33] The JPO does not use the ECLA. However, it uses its own system, the F-term, which, like the ECLA, provides greater detail than the IPC. F-terms consist of a 5-digit theme code, a 2-digit viewpoint symbol, and a 2-digit number. For example, consider technologies classified as 4D002 AA12. The 5-digit theme code 4D002 pertains to "processing of waste gases." The two letter viewpoint symbol AA pertains to "compounds to be processed." Finally, the 2-digit number specifies that NO_X is the component processed. By searching by year, researchers can get lists of patents for a given application year. However, the priority date of these patents is not available, nor is it possible to separately identify foreign and domestic Japanese patents using this database.

---

[32] Available at http://www.delphion.com.

[33] This database can be found at http://www.ipdl.jpo.go.jp/homepg_e.ipdl.