Induced Innovation and International Environmental Agreements: Evidence from the Ozone Regime*

Eugenie Dugoua[†]

January 14, 2020

Abstract

Can international environmental agreements induce innovation on green technologies? Noncooperative game theory suggests that agreements only occur when costs to the players are low. This likely happens once technological solutions are available. Agreements then may help diffuse technologies rather than fostering their development. I provide the first quantitative evidence that the Montreal Protocol to protect the ozone layer triggered a large increase in research and innovation on alternatives to ozone-depleting molecules. To do this, I use a DiD design to compare innovation and scientific output related to different molecules. I leverage the full text of patents and scientific articles and apply machine learning-based methods to construct molecule-level chemical and industrial characteristics. My results challenge the view that agreements foster technological diffusion without affecting much of the dynamics of innovation and bear critical implications for how we think about environmental agreements.

JEL Classification: O31, O33, Q55, Q58, F53
Keywords: induced innovation, directed technological change, international environmental agreements, ozone, CFCs, Montreal Protocol, climate change, difference-in-differences, synthetic control, topic modeling

Like national environmental policies, international environmental agreements attempt to mitigate environmental degradation. However, when problems run across national borders (like ozone depletion or climate change), governments usually lack incentives to create domestic policies. International agreements, then, are needed. Ozone depletion is hailed as the greatest environmental success, but we often hear that the agreement did not really matter to solve it. In fact, noncooperative game theory suggests that agreements only occur when costs to the players are low. In other words, agreements occur only once technological solutions are readily available, and they simply contribute to diffusing these technologies, as opposed to fostering the development of new ones.

This paper shows that, on the contrary, agreements can induce innovation and that agreements, therefore, are part of the process of delivering cheaper environmental-friendly technologies. This provides a strong argument for negotiating ambitious agreements as early as possible since technologies are too often the keystone for addressing environmental problems. To make this argument, I provide empirical evidence from the Montreal Protocol and its following amendments. In 1987, at Montreal, high-income countries decided to phase-out chlorofluorocarbons (CFCs) from industrial activities because CFCs were known to destroy the protective layer of ozone molecules in the stratosphere. Technological change unrolled rapidly: within a decade, the production and consumption of CFCs decreased by more than 80%[1]. The protocol is still hailed as one the most successful environmental international agreement and remains a point of reference in policy discussions about global environmental problems such as climate change.

Despite the large scholarly literature on the topic, the dynamics of innovation in the ozone crisis are still debated. Richard E. Benedick, chief U.S. negotiator at Montreal, claims the agreement triggered a vast effort in research to find CFC substitutes (Benedick 2009). But others emphasize that CFC substitutes were already available at the time of negotiations (Heal 2016; Sunstein 2007). This paper is not only the first to empirically show that Montreal fostered innovation, but it also quantifies its effect. I do this using a novel molecule-level panel dataset with both a difference-in-differences (DiD) strategy and a synthetic control method (SCM). Additionally, I apply machine

---

[1] My calculations using UNEP data.

learning methods to semantically match documents and measure similarity between molecules.

Developing CFC substitutes did not, in fact, require scientists to identify and create new chemical structures. The set of molecules with greatest potential to be CFC substitutes was already well-known at the time. I compile a list of 14 of such molecules and consider those molecules as treated by the signature of the Montreal Protocol. The technological challenge lied, instead, in finding out how such molecules could be used in the myriad of industrial processes that required CFCs, cost-effectively and at large scale. This meant scientists had to learn about their thermodynamics properties, toxicity profile and environmental acceptability. To capture this research effort, I collect scientific articles published in journals indexed by Science Direct between 1970 and 2000. Firms also had to develop new processes and formula designs to retrofit installed equipment with the CFC substitutes or to altogether replace them. To track progress on these aspects, I collect patents granted by the United States Patent and Trade Office (USPTO) between 1976 and 2000.

Unfortunately, no preexisting classifications allow me to easily identify which documents relate to CFC substitutes. I, therefore, search the full text of patents and articles for mentions of any of the 14 CFC substitutes, and construct a panel dataset where each observation is the number of documents mentioning a given molecule at least once in a given year. Finally, I apply machine learning-based text analysis techniques to construct variables that proxy the scientific and industrial context of the molecules. These variables correspond to the proportion of specific topics present in documents mentioning molecule i. Intuitively, they describe the type of words associated with molecule i. I begin by estimating the difference before and after the signing of Montreal in the number of documents mentioning CFC substitutes. As illustrated on Figure 1, I find increases larger than 500% for both patents and articles.

To account for potential underlying trends, I compare innovation on CFC substitutes with a control group of molecules known as hazardous air pollutants (HAPs). I argue that these molecules can serve as controls because they are unrelated to ozone or CFCs and, just like CFC substitutes, are used in a diverse range of industrial applications. Importantly, pre-trends in the number of documents mentioning both sets of molecules are comparable. The preferred DiD specification
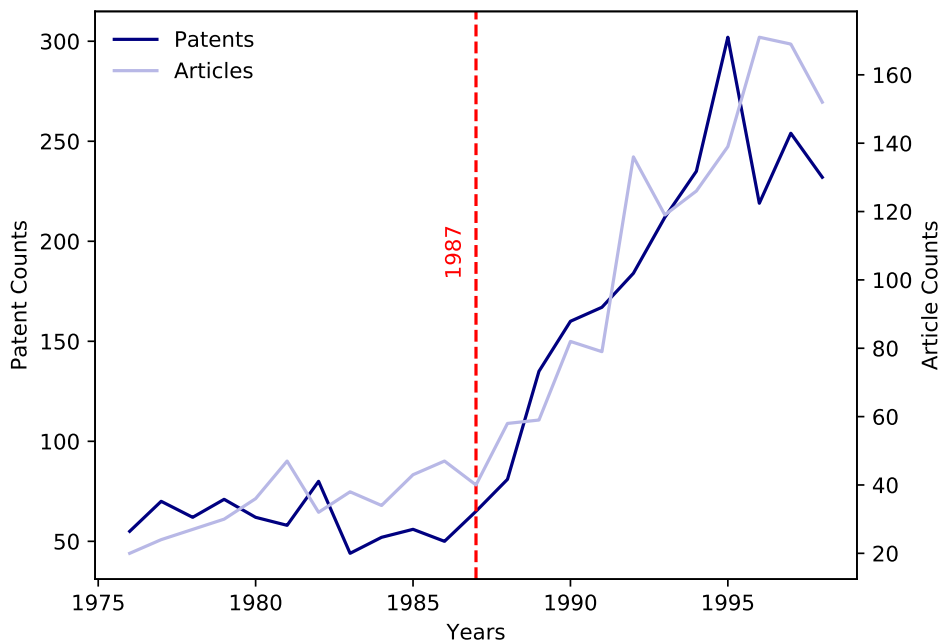
Figure 1: Counts of patents and articles mentioning CFC substitutes

Note: The graph plots the yearly number of patents and articles mentioning the names of any of the 14 CFC substitutes. For patents, the graph shows any patent granted between 1976 and 1999. The year on the x-axis, however, corresponds to the application date. For articles, the year on the x-axis corresponds to the year the article was published in the academic journal. We note a clear increase for both patents and articles after 1987, the year Montreal was signed.

controls for topic proportions to help account for heterogeneity along chemical and industrial dimensions. Results indicate that Montreal led to an increase close to 4000% for patents and larger than 500% for articles. This corresponds to average increases of about 17 patents and 8 articles per year.

The increase in the number of documents becomes statistically significant as from 1989 for patents and 1990 for articles, two to three years after the agreement was signed. Such lag is not unlike other findings in the literature (see for example Popp (2002)) and can be attributed to the necessary delays involved in turning research efforts into patent applications and published academic papers. The estimates are robust to a series of alternative specifications, including weigthing counts by the number of times the molecules are mentioned in each document, as well as weighting by the number of citations that documents received.

Patents and articles often mention several CFC substitutes in the same document, and the observations used in the DiD design (14 CFC substitutes) are therefore not independent. Another

approach consists in considering them in aggregate, as one treated molecule. To estimate a treatment effect on such "aggregate CFC substitute", I use a synthetic control method (SCM) (Abadie et al. 2010, 2015) where the weights are chosen so that the synthetic control resembles the treated unit along topic proportions. SCM results are consistent with the DiD and estimates an increase of about 135% in the number of patents, and around a 180% increase in the number of articles.

This paper contributes to the literature on technological change and the environment (Jaffe et al. 2002; Popp 2019; Popp et al. 2010). Specifically, it complements the strand of empirical research examining how green innovations can be induced. For example, by increases in energy prices (Aghion et al. 2016; Popp 2002) or by domestic environmental regulations (e.g., Calel et al. (2016) and Jaffe et al. (1997)). To my knowledge, only one other paper looks at the causal effect of an international environmental agreement on technological change but the authors focus on transboundary local air pollution (Dekker et al. 2012) whereas I focus on a case of a global public good (strastospheric ozone)[2].

While scholars have thoroughly investigated the game theoretic and diplomatic aspects of the ozone crisis (Barrett 2003; Benedick 2009; Murdoch et al. 2009; Parson 2003; Wagner 2009), no quantitative analysis of the dynamics of innovation during the crisis has been carried out. This is despite the economics, science, and politics of ozone serving as an anchor point for our understanding and beliefs about the role of diplomacy, agreements, and technologies in solving environmental issues, especially climate change (Barrett 1999; Sunstein 2007). This paper thus complements the literature on Montreal by showing and quantifying its effect on science and innovation.

I start with providing some background information in Section 1. I then describe data sources in Section 2, the empirical strategy in Section 3 and results in Section 4. Section 5 considers and discusses alternative hypotheses and mechanisms. Finally, in Section 6, I discuss how my results connect to the theoretical literature on international environmental agreement.

---

[2]Several studies investigate the effect of international environmental agreements on pollution outcomes (Aichele et al. 2011; Finus et al. 2003; Kellenberg et al. 2014) but they seldom look at the impact on science and innovation.

1 Background

1.1 Theory and Hypotheses

The concept of *induced innovation* goes back to Hicks (1932) who stated that innovations are biased towards high priced factors to make their use more efficient or to substitute them[3]. The theory was easily extended to environmental policies. In the simplest model, environmental regulations force firms to allocate some inputs (labor or capital) to pollution reduction, or restrict the choice of technologies and inputs in the production process (Milliman et al. 1989). Environmental regulations are thought, therefore, as extra costs hurting firms' profitability. But, doing so, they induce firms to innovate in order to offset the costs (at least partially, and according to the *Porter Hypothesis*, sometimes even more than fully (Ambec et al. 2013; Porter et al. 1995)).

The theory can easily be extended to international environmental agreements. If the signature of an agreement does not immediatly force firms to adapt their production process[4], it nonetheless changes their expectations of future domestic environmental policies. At least if the agreement is perceived as binding, and, arguably, this is what the Montreal Protocol did. Montreal included trade restrictions with non-parties in ozone-depleting substances (as well as in products containing those substances). It also included the threat of banning trade in products made using ozone-depleting substances.

Barrett (1999) suggested Montreal's trade measures solved the enforcement problem, and Wagner (2016) further argued they promoted full participation in the protocol, ensuring its almost-universal ratification. Hence, Montreal's trade restrictions effectively acted as a mechanism for free-rider deterrence and leakage prevention rendering the agreement binding. The main hypothesis, therefore, is that Montreal credibly signaled firms that the continued use of CFCs would become increasingly costly, and induce them to innovate. Richard Benedick, the U.S. head negotiator at Montreal, argued that "(it) was evident (...) that the protocol was in fact moving industry in directions that two years earlier had been considered impossible." (Benedick 2009,

---

[3]In the past two decades, the concept has reappeared under the phrase "directed technical change" (Acemoglu 1998) encompassing not just price effects, but also market size and regulatory effects.

[4]This would happen only once the agreement is implemented into domestic policy

Chap.8 p.104.). But Benedick only refers to articles published in the *New York Times* and *Chemical and Engineering News* to support his claim.

On the other hand, an often heard narrative argues that CFC substitutes were readily available before the negotiations. For example, Sunstein (2007) claims that "an international agreement was largely in the interest of American manufacturers, which had already initiated a transition to safe CFC-alternatives". This view is also often expressed in media outlets. For example, *The New York Times* on August 20, 2002 stated that "(the) agreement's success occurred, in large part, because substitutes for the harmful chemicals were readily available (...)." Importantly, such claims are consistent with the theory of why and when agreements appear (Barrett 1994). In the case of Montreal, that theory implies that Montreal was successfully negotiated because the costs of doing so were low relative to benefits.

In this paper, I rely on quantitative analysis in the hope to disantangle these different hypotheses. I discuss how to reconciliate the empirical results with the theory of international environmental agreements in Section 6.

1.2 Events that led to Montreal

The story of the ozone crisis began in 1974 when two chemists laid out the theoretical possibility that CFCs broke down ozone molecules in the stratosphere (Molina et al. 1974). The harmful effects of a thinner ozone layer were not well understood, but it was clear that more UV light would cause more skin cancers, eye cataracts and, likely, lower productivity in fishery and agriculture (Miller et al. 1986). Consequently, in the late 1970s, the issue began to take prominence in the media and regulators' minds.

A few countries and firms unilaterally decided to take action. In August 1977, the U.S. Congress wrote into law a CFC ban on aerosols by 1978[5], and firms such as DuPont removed CFCs from their spray products because they worried about their public image. These pre-Montreal domestic regulations and corporate decisions can be viewed as zero-cost unilateral moves. They targeted a very specific industrial application of CFCs for which substitutes could easily and cheaply be

---

[5]Similarly, in 1978, Canada, Switzerland and Scandinavian countries all banned CFC aerosols. On continental Europe, Germany called for a European Community-wide ban, but without success.

implemented[67]. They, therefore, did not require a significant research effort. The low levels in patent and article count between 1970 and 1987 on Figure 1 indicates that neither aerosol regulations nor consumer pressure seemed to have stimulated science and innovation on the 14 CFC substitutes I consider[8].

In the beginning of the 1980s, the issue stopped being a priority. Uncertainties in the science of atmospheric ozone seemed irreducible, and the year 1981 saw the election of a strongly anti-regulatory American administration. In Europe as well, many governments persisted in refusing to harm their domestic manufacturers with any regulation. Edward A. Parson highlights that, although some manufacturers initially started research on potential substitutes in the late 1970s, these efforts quickly came to an end around 1981 (Parson 2003, Chap.3 p.53 and Chap.7 p.173). Parson argues that manufacturers stopped these R&D programs because they had determined that CFC substitutes would cost around two to five times more than CFCs, and it made no sense to continue developing these substitutes with little sign of regulations under way.

Benedick emphasizes the complexity of the negotiations and the great uncertainty of the negotiations' outcome. He further argues that some exceptional turns of events unlocked the situation. Unexpectedly, Reagan overruled his own administration and approved the agreement: the U.S. President had a skin cancer removed twice in the past, and Benedick hints that Reagan's life experiences weighed heavily on his decision. On the European side, the biggest opponent to the regulation of CFCs, the U.K., left the European Community Presidency, leaving Germany, Denmark, and Belgium, firm proponents, as the head negotiators. This account of the negotiations' success does indeed indicate that the agreement largely occurred independently from the state of R&D activities on CFC substitutes.

1.3 Which molecules were "treated"?

CFCs have a very specific molecular structure: they contain only carbon, chlorine and fluorine atoms. Such chemical structure is what conferred CFCs attractive thermodynamics properties:

---

[6]Physical substitutes included roll-on devices; chemical substitutes included alkanes.

[7]In 1980, the EPA proposed to freeze other uses beyond aerosols but U.S. industry blocked the initiative.

[8]These 14 CFC substitutes were targeting foams, refrigeration and solvent applications of CFC.

they are unusually stable, nonflammable, nontoxic and noncorrosive. In a word, they were ideal for manufacturing many consumer goods[9]. Best of all, they were cheap to produce. Consequently, they became broadly used in many different industries such as foams, refrigeration and air-conditioning, aerosols, fire protection and solvents[10].

Strategies for reducing CFCs included physical substitutes (like pump-action sprays instead of aerosol sprays) or recycling. But the most critical applications (e.g. air-conditioning units) needed chemical substitutes. Importantly, the intricate relationship between molecular structure and industrial properties implied that there wasn't an infinite number of possible substitutes: good candidates would present similar alkane chains where hydrogen atoms replace halogens. For example, and as detailed on Figure 2, when a chlorine atom in CFC-12 is substituted by a hydrogen, we get HCFC-22. If, instead, it is substituted by a methyl group, we obtain HCFC-142b. Here "HCFC" stands for hydro-chlorofluorocarbons. When all chlorine atoms are substituted with hydrogens, the compounds are then known as HFC, or hydro-fluorocarbonsfootnote.

It was public knowledge, therefore, that the quest for CFCs subtitutes lied in the realm of HCFCs and HFCs. These molecules had been known for a long time, at least on paper and in the lab[11]. Hence, developing CFC substitutes was not so much about "new-to-the-world" compounds, but instead "new-to-the-industry" compounds. The key technological challenges lied in making large-scale production cost-efficient, redesigning processes and equipment already installed, and learning about environmental acceptability and human toxicity.

To conduct my empirical analysis, I construct a list of potential substitutes using historical records. After Montreal, manufacturers from the US, Europe, and Japan received authorization from antitrust officials to organize cooperation, at least on the science for which patenting was not possible. They launched two working groups to study the feasibility of various alternatives. The

---

[9]Initially, CFCs somewhat embodied the miracle of modern chemistry. They were first commercially used in 1928 as cooling fluids for refrigerators, and were specifically designed to substitute other dangerous refrigerants that were either toxic or inflammable (Parson 2003).

[10]CFCs are great refrigerants because they vaporize at low temperature and are very energy-efficient coolants. As aerosols, they were used in cosmetics, household products, pharmaceuticals, and cleaners. Finally, their nonreactive property made them key products for cleaning microchips and telecommunication equipment

[11]The first-ever granted patents related to HCFCs and HFCs typically go back to the 1930s; at the time, chemists were experimenting with halogenation processes and heat transfers. For example, in 1934, a patent is claimed for a "method of producing refrigeration which comprises evaporating in the vicinity of a body to be cooled and subsequently condensing CH_2ClF." US Patent 1,968,049. CH_2ClF is a.k.a. HCFC-22.

PAFT (Program for Alternative Fluorocarbon Toxicity Testing), created in January 1988, worked on assessing the toxicity of five possible alternatives. The AFEAS (Alternative Fluorocarbon Environmental Acceptability Study), created in December 1988, investigated the atmospheric dynamics of twelve potential CFC substitutes. Importantly, none of these reports focused on industrial aspects of the molecules but rather how they would interact with human helath and the environment once released in the atmosphere. I use these twelve molecules and also include two other possible CFC substitutes mentioned in Benedick (2009) and Parson (2003)[12].

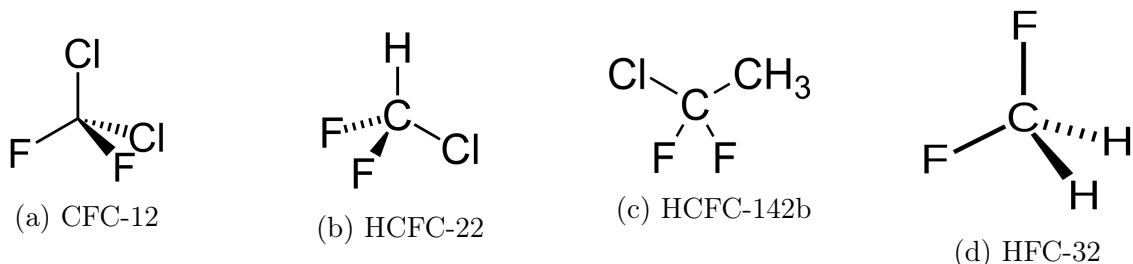(a) CFC-12   (b) HCFC-22   (c) HCFC-142b   (d) HFC-32

Figure 2: Molecular Structure of CFCs, HCFCs and HFCs

Note: CFC stands for chlorofluorocarbon, i.e. a molecule entirely made of carbon, chlorine and florine atoms. When one chlorine atom in CFC-12 is substituted with a hydrogen, it becomes HCFC-22, but if substituted with a methyl group, we obtain HCFC-142b. Here "HCFC" stands for hydro-chlorofluorocarbons. When all chlorine atoms are substituted with hydrogens, the compounds are then known hydro-fluorocarbons, or HFCs. For example, when the two chlorine atoms in CFC-12 are replaced by hydrogens, we get HFC-32.

2 Data Description

2.1 Patents

Patent data has been used extensively in empirical research in the past two decades[13], and I follow this literature by using patent counts as a proxy for innovation. Specifically, I use the full-text of U.S. patent grants published between 1976 and 2000 [14]. We know from surveys that, in the chemical industry, patenting is an important way of protecting competitive advantage from new products and processes (Sampat 2018). Since chemicals can often easily be "reverse

---

[12]HFC-245fa and HFC-365mfc are mentioned as possible substitutes in foams. The supporting online material shows the name and additional information about all molecules considered in the analysis.

[13]Examples and reviews include Hall et al. (2012), Henderson et al. (1998), Popp (2005), and Williams (2013, 2017).

[14]I download the patent data from the U.S.P.T.O. repository. It represents a total of 2,605,925 patents.

engineered"[15], secrecy offers a limited mode of appropriation. Strong incentives exist, therefore, to use patenting either to protect inventions from being copying or to prevent competitors to patent related inventions (Cohen et al. 2000).

I use the texts contained in the abstract and summary description of the invention as they are likely to faithfully represent the invention. The cleaning procedure involves a series of standard steps that are detailed in Appendix A. Patents contain the names and addresses of inventors and assignees[16]. I categorize patents by the type of their assignees' affiliation (e.g., business, education, or government). More details about how the meta-data is cleaned, matched and classified by type are provided in Appendix A. I use the OECD Citations database File[17] to obtain data on the number of forward citations received by each patent. Finally, I sort patents by application date (as opposed to the date when they are granted).

2.2 Articles

Similar to patents, scientific articles have been used extensively in empirical research in the past two decades[18], and I follow this literature by using articles counts as a proxy for increases in knowledge. Specifically, I collect scientific articles published between 1970 and 2000 in journals indexed by ScienceDirect[19]. Although not the complete universe of scientific articles, the sample provides a good coverage of journals in fields related to engineering and physical sciences, which will allow me to characterize trends in the emergence of scientific knowledge related to CFCs substitutes.

To download the full-text of articles, I first collect each journal's ISSN number[20], and then use

---

[15]See for example Moser (2012) for an examination of how the publication of the periodic table in 1869 made chemicals easier to reverse engineer and led chemical inventors to shift from secrecy to patents in the mid-nineteenth century.

[16]To associate patents to specific countries, I use the country of the assignee. When patents have no assignee but only inventors, I use the country of the inventor.

[17]OECD, Citations database, February 2019

[18]Examples and reviews include Azoulay et al. (2010), Azoulay et al. (2019), Iaria et al. (2018), Redner (2004), Thompson et al. (2005), and Wang et al. (2013).

[19]ScienceDirect hosts articles from about 2,500 academic journals published by Elsevier (`https://www.elsevier.com/solutions/sciencedirect`).

[20]These are listed on Elsevier's website for specific disciplines: `https://www.elsevier.com/solutions/sciencedirect/content/journal-title-lists`. I focus on the following disciplines: chemistry, chemical engineering, engineering, environmental science, materials science, and physics and astronomy.

ScienceDirect's API to obtain the list of identifiers for articles published in each journal. I then query ScienceDirect full-text's API with each identifier[21]. After a series of cleaning procedures, I obtain a total number of 1,811,301 articles. I proceed with a series of cleaning steps as described in Appendix A. For data on affiliations and citation counts, I query the Scopus search API, and use the Global Research Identifier Database[22] to classify authors' affiliations (e.g., education, company, etc...).

2.3   Searching for Molecule Names

I search through every patent and article to find documents containing the name of the molecules of interest. But chemical compounds are often given several names; for example, HCFC-22 has 39 other possible names such as chlorodifluoromethane or algeon 22. I develop an automatic script to collect all possible names on SciFinder, a database of chemical information maintained by the American Chemical Society[23]. I then search through all patents and articles to identify the documents in which any of the names appear[24]. This allow me to confidently capture all occurrences of a molecule. I proceed similarly to identify the patents and articles that mention any of the 171 HAPs. In total, I find 3270 patents and 1926 articles mentioning at least one CFC substitute. The search procedure, however, may lead to measurement error with over-detection if patents and articles mention molecules despite these same molecules not being the core subject of the document. There is no reason to believe that over-detection will be greater for CFC substitutes than for the control group. Hence I don't expect over-detection to bias the results. I nonetheless run several robustness checks with respect to the number of occurrences of molecules in Section 4.2.

---

[21]Full text data was successfully downloaded for 1,843,684 articles, out of a total of 2,307,345 DOIs initially returned by the API. This implies that Elsevier listed 463,661 DOIs for which the full text was not available. This might be due, for example, to entire journals dropping out of Elsevier Science Direct's collection.

[22]https://www.grid.ac/

[23]A full list of all the possible names of CFC substitutes is shown in the supporting online material.

[24]I look for any English name listed in SciFinder but I do not look for chemical symbols. The articles' text is usually the output of optical character recognition, and chemical symbols and formulae are too often rendered with mistakes.

2.4 Descriptive Statistics on Documents Mentioning Substitutes

Tables 1, 2 and 3 illustrate the types of patents and articles found to mention CFC subtitutes. The most common patent codes are related to chemical compounds containing halogen atoms and the most cited patents corresponds to CFC subtitutes innovations for the pharmaceutical sector. The most cited articles, on the other hand, seem to focus on physical chacteristics or new synthesis routes. Table D1 n Appendix displays summary statistics about countries and affiliations of patent assignees and authors of articles. The typical patent is granted to a for-profit organization in the United States. Specifically, more than 96% of patents are granted to for-profit organizations while the rest is shared among organizations coming from the educational and public sectors. European assignees represent around 25% of patents; Japanese around 12%.

Table 1: Five most common patent codes for patents mentioning CFC substitutes

ICL	Count	Description
C07C	357	Acyclic or carbocyclic compounds
C08J	156	General processes of compounding
C09K	147	Materials for applications not otherwise provided for
C08G	84	Compounds of unknown constitution
C10M	73	Lubricating compositions

Note: The table displays the most frequent codes associated to patents mentioning CFC substitutes. As expected, most codes belong to the C class ("Chemistry, Metallurgy"). The subclasses "C07" and "C08" refer to the preparation (e.g., purification, separation or stabilisation) of organic compounds. As such, they encompass any patent related to compounds containing carbon and halogens atoms (e.g., C07C 19/00: Acyclic saturated compounds containing halogen atoms). To limit noise, the sample used to generate the table contains only documents with at least three occurrences of CFC substitutes.

Table 2: Titles of the five most cited patents mentioning CFC substitutes

Nbr Cit	Year	Assignee	Title
104	1995	Glaxo Group Limited, UK	Aerosol formulations containing P134a and salbutamol
103	1995	Glaxo Group Limited, UK	Aerosol formulations containing P134a and particulate medicaments
101	1995	Glaxo Group Limited, UK	Aerosol formulations containing propellant 134a and fluticasone
97	1995	Riker Laboratories, Inc., USA	Medicinal aerosol formulations

Note: The table displays the titles of the most cited patents mentioning CFC substitutes. Patent citation patterns vary greatly across industries. The fact that the most cited patents here all relate to pharmaceuticals applications (e.g. aerosol formulation of a drug) may only be indicative of that sector's greater patenting output or tendency to cite more. To limit noise, the sample used to generate the table contains only documents with at least three occurrences of CFC substitutes.

Table 3: Titles of the five most cited articles mentioning CFC substitutes

Nbr Cit	Year	Title	Journal	Affiliation 1st author
509	1992	Organic peroxy radicals: Kinetics, spectroscopy and tropospheric chemistry	Atmospheric Environment Part A	Academia (DE, UK, FR)
419	1982	Evaporative heat transfer, pressure drop and critical heat flux in a small vertical tube with R-113	International Journal of Heat and Mass Transfer	GE Global Research (USA)
401	1992	Environmental catalysis	Applied Catalysis B: Environmental	Air Products & Chem. Inc (USA)
346	1993	Synthesis of chiral and bioactive fluoroorganic compounds	Tetrahedron	Academia (IT)
333	1996	Methods for the synthesis of gem-difluoromethylene compounds	Tetrahedron	James Black Foundation (UK)

Note: The table displays the titles of the most cited articles mentioning CFC substitutes. As expected, articles focus on the chemical and physical characteristics of CFC substitutes ("kinetics,", "evaporative heat transfer", etc...) as well as on synthesis routes. To limit noise, the sample used to generate the table contains only documents with at least three occurrences of CFC substitutes.

3 Empirical Strategy

3.1 A Sharp Post-1987 Increase

Next, I document the temporal trends in the yearly count of documents about CFC substitutes and test whether there is a change of patterns before and after the date of the signature of the agreement. As Figure 1 illustrates, the numbers of patents and articles increase after Montreal was signed in 1987[25]. To quantitatively investigate the temporal patterns, I implement a first difference specification with a mean shift (Equation 1) and a first difference specification with a trend-break (Equation 2). $Count_{m,t}$ is the number of documents in year t mentioning molecule m; $\lambda_{post1987}$ is a dummy variable that equals one when $t > 1987$; λ_m are molecule fixed effects; Y is a continuous variable indicating the number of years relative to 1987. The sample here consists of the 14 different CFC substitutes for which I track the number of patents and articles throughout the years. I run separate regressions for patents and articles, and bootstrap standard errors.

$$\text{Count}_{mt} = \alpha + \beta_0 \times \lambda_{post1987} + \lambda_m + \epsilon_{mt} \tag{1}$$

---

[25]Figure B1 in Appendix B plots time-series similar to Figure 1, but each CFC substitutes separately. The post-1987 increase is present for most substitutes.

$$\text{Count}_{mt} = \alpha + \beta_1 \times Y \times \lambda_{post1987} + \beta_2 \times Y + \lambda_m + \epsilon_{mt} \tag{2}$$

The main hypothesis is that β_0 and β_1 are both positive for CFC substitutes, implying significant increases in research and patenting activities relating to CFC substitutes after 1987 once Montreal passed. Table 4 presents the results for these specifications. Model 1 confirms that there is a significant and positive mean shift after 1987 in the number of patents and articles mentioning CFC substitutes. The coefficients indicate almost 30 additional patents and around 13 additional articles for the average CFC substitute every year after 1987. This corresponds to a 551% increase (594% increase) in the number of patents (articles) for the years 1987-2000 relative to the years 1975-1986. Model 2 shows that the change can also be modeled as a trend break. The coefficient for "Years" indicates that there is a small positive underlying trend for articles.

Table 4: First differences for CFC substitutes

	(1) Patents	(2) Patents	(3) Articles	(4) Articles
Post 1987	29.510***	6.097**	13.022***	2.113
	(2.113)	(2.632)	(1.072)	(1.581)
Post 1987 x Years		3.955***		1.442***
		(0.439)		(0.284)
Years		-0.027		0.156**
		(0.246)		(0.063)
Molecule FE	Yes	Yes	Yes	Yes
Bootstraped	Yes	Yes	Yes	Yes
R-squared	0.639	0.736	0.577	0.630
Observations	322	322	406	406

Standard errors in parentheses
Years are relative to 1987.
Time-series: (1976-1998) for patents; (1970-1998) for articles
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The tables present regression results for first difference specifications. Model 1 confirms that there is a significant and positive mean shift after 1987 in the number of patents and articles mentioning CFC substitutes. Model 2 indicates that the change can also be modeled as a trend break. The coefficient for 'Years' indicates that there is a small but statistically significant positive underlying trend for articles.

At this stage, one shall not exclude that the observed post-1987 increase may be due to other temporal confounding factors (e.g. policies or macroeconomic conditions fostering academic and industrial research in the 1990s). To further investigate whether the post-1987 increase can be causally linked to Montreal, I use a set of molecules known as HAPs as a comparison group.

3.2 HAPs as a Comparison Group

The challenge in choosing a group of control molecules resides in finding molecules as similar as possible to the treated molecules, while, at the same time, not affected by treatment. HAPs is an umbrella term for molecules categorized as hazardous air pollutants and which became monitored under the 1990 Clean Air Act due to human health concerns including cancer, asthma, birth defects, reproductive effects, and neurodevelopmental effects, as well as adverse ecological impacts. Examples include benzene, chromium or formaldehyde[26].
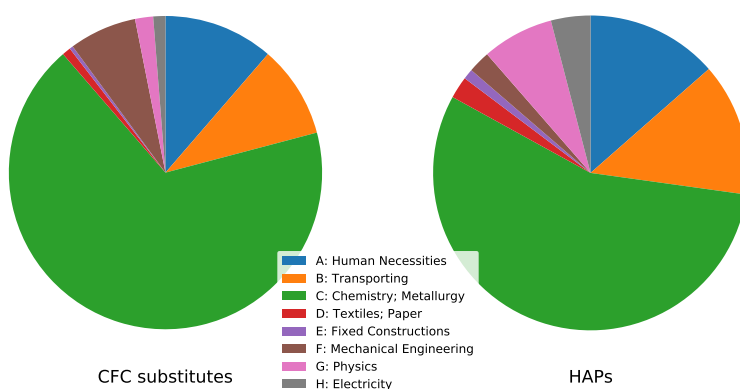

Figure 3: Top level patent codes for CFC substitutes and HAPs

Note: The figure shows that, overall, patents mentioning CFC substitutes and HAPs fall into similar top-level codes. HAPs are a group of 171 molecules that have no relationship to ozone and that are used for diverse industrial applications. The figure indicates the two groups of molecules present important similarities which motivates the use of HAPs as control molecules to estimate the causal effect of the post-Montreal regime. The patent codes are from the international patent classification.

Two reasons motivate the use of HAPs as a comparison group for CFC substitutes. First, HAPs are related to industrial activities similar those of CFC substitutes. Figure 3 illustrates that, overall, patents about CFC substitutes and HAPs fall into similar top-level codes. Figure B2 further shows that they also display similar second-level patent codes. Second, the HAPs I consider in my analysis have no connection to ozone depletion[27], and none of the ozone-depleting substances were ever categorized as HAPs.

A concern is that the Clean Air Act may have impacted research and innovation related to

---

[26]The full list of HAPs is displayed in the supporting online material.

[27]Three HAPs (chlorine, methylenechloride, trichloroethylene) were suggested as possible CFC substitutes, but I exclude them from the sample. I use the EPA SNAPs website to obtain the list of compounds or devices suggested by firms to the EPA as possible alternatives to CFCs.

HAPs. In 1990, an amendment to the Clean Air Act required the EPA to promulgate regulations establishing emission standards for major sources of HAPs. But, in practice, the EPA published the initial promulgation schedule only in 1993[28]. I take a conservative approach and limit my analysis to the time-period until 1992 for patents and 1995 for articles. I allow for an additional three-year period for articles since the timestamp for articles is the date of publication while the timestamp for patents is the application date.

As I use HAPs to construct a counterfactual, the treatment should be understood broadly as the "ozone regime". As such, it includes the signature of the Protocol in Montreal in September 1987, the following country-by-country ratifications[29], and the following amendments to the Protocol (the London and Copenhagen revisions in 1990 and 1992 respectively). It also includes, therefore, the implementation of these international treaties into domestic regulations. In the US for example, this was done through the Clean Air Act amendment of 1990. The counterfactual therefore represents a world without any of those interventions, and in particular a world without, or with limited, unilateral actions. The inherent challenges of any global public good problem provide a good case for arguing that a world with limited unilateral actions is an apropriate counterfactual. The absence of costly unilateral actions before 1987 is further testimony to those challenges[30].

Although overall HAPs are related to industrial activities similar to those of CFC substitutes, some HAPs will be more similar to CFC substitutes than others. In particular, as Figure 4 illustrates[31], some have much higher counts than the average CFC substitute. In what follows, I adopt a strategy where only HAPs with similar counts are included in the control group[32]. Going further still, one would like to ensure that the HAPs chosen are similar to CFC substitutes not just in terms of counts but also in terms of chemical, physical and industrial properties. To do so, I leverage topic modeling.

---

[28]The EPA published the initial list of "source categories" in 1992 (i.e. the list of industries and production processes targeted by the regulations), and in 1993, the promulgation schedule specified by which year sectors were expected to comply with the emission standards for each category or subcategory of major sources and area sources of HAPs. This was known as the National Emission Standards for Hazardous Air Pollutant (NESHAP), and most sectors were asked to comply by 1997 or 2000.

[29]The USA ratified in April 1988; European countries in December 1988.

[30]The aerosol bans in the 70s and 80s were not costly because physical and chemical substitutes existed; for example, roll-on deodorants instead of spray deodorants.

[31]A similar graph can be found for articles in Appendix B Figure B3

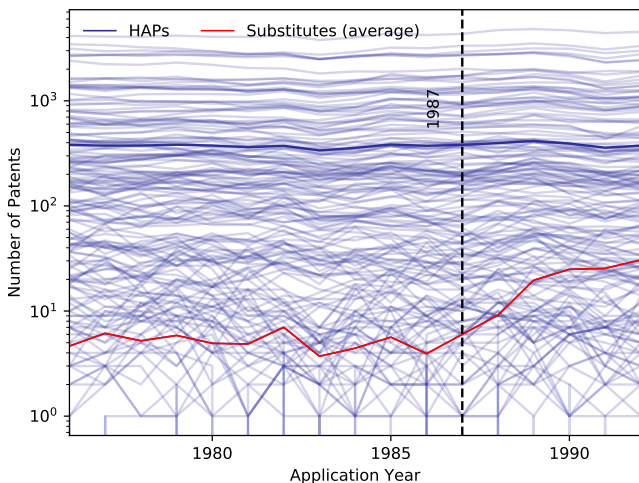[32]I provide full details on how I construct the control group in Section 4.

Figure 4: Patent counts for each HAP and for the average CFC substitute

Note: The grap shows patent counts for each HAP (thin lines), for HAPs on average (thick line labeled "HAPs') and for CFC substitutes on average. The graph illustrates that HAPs are a diverse group of molecules. In particular, some of them have counts much higher than the average CFC substitute.

3.3 Topic Modelling

I use topic modeling, a machine learning method for text analysis, to generate covariates that describe the semantics of the documents mentioning molecule names. These covariates, therefore, help describe and measure molecules' chemical and industrial characteristics. In practice, the experimenter chooses a number of topics, and after training the algorithm on a corpus, the procedure outputs document-level *topic proportions*, that is a variable from 0 to 1 indicating to what extend topic i is present in a particular document. Put differently, using the words that appear in a given document; the LDA model infers what proportion of each topic a document contains.

Intuitively, the topic proportions describe quantitatively what an article talks about, and we can, therefore, think of it as a proxy of the physical, chemical and industrial characteristics of a molecule. Specifically, I run several LDA models with different number of topics (Blei 2012; Blei et al. 2009), and compute a coherence score for each of them (Röder et al. 2015)[33]. For each corpus, I choose the lowest number of topics that seems to offer one of the highest coherence score: 20 for patents and 15 for articles[34].

---

[33]Figure A2 in Appendix A shows that coherence increases with the number of topics up to a certain point.

[34]I train the algorithm, not on the entire corpus, but on the subset of documents that contain at least one mention of a molecule

Table 5 provides a few examples of words in some topics[35]. For example, for patents, the most likely word in Topic 1 is "polymer" (with probability 0.0161). Then come "catalyst" and "carbon". This suggests that Topic 1 could be labeled as "Catalysts on polymer substrates"[36]. However, the documents being highly technical in nature, topics can be difficult to interpret.

Once topics are discovered, I obtain topic proportions for each document and aggregate them at the molecule level. To do so, I calculate weighted means where the weights are proportional to the number of times a document mentions a molecule[37]. Figure A4 displays a series a scatterplots illustrating the variation in topic proportions across molecules. Some HAPs have values of topic proportions that stand out as outliers, indicating that they present semantic contexts that are very different from the one of CFC substitutes. This, in essence, illustrates why topic proportions are useful: they will allow to control for variation in how different molecules are mentioned in the text of documents.

---

[35]Table A1 and A2 in Appendix A provides the full list of words and topics.

[36]Catalysts are molecules (typically "metals", word 7 in topic 1) used to start and maintain chemical reactions. They are often made more effective by being "attached" to a polymer substrate. Hence, the production of such catalysts often involves a "polymerization" process (word 6 in topic 1) where free "radicals" (word 13 in topic 1) ensure the addition of new monomers to the polymer chain.

[37]Figure A1 in Appendix B summarizes these steps.

Table 5: Top five words in selected topics

	(a) Patents				(b) Articles	
	Words	Probability			Words	Probability
Topic 1	polymer	0.0161		Topic 1	compound	0.0162
	catalyst	0.0123			extract	0.0072
	carbon	0.0095			structure	0.0068
	weight	0.0094			product	0.0061
	atom	0.0087			methyl	0.0056
Topic 2	metal	0.0084		Topic 2	surface	0.0155
	membrane	0.0082			material	0.0096
	solution	0.0071			layer	0.0086
	particle	0.0067			film	0.0075
	surface	0.0065			process	0.0062
Topic 3	formula	0.0118		Topic 3	laser	0.0129
	carbon	0.0092			signal	0.0102
	atom	0.0088			sample	0.0097
	substitute	0.0086			pulse	0.0092
	amine	0.0077			radical	0.0081
Topic 4	agent	0.0147		Topic 4	gifhttps	0.0351
	composition	0.0112			thumbnail	0.0282
	active	0.0064			downsample	0.0270
	weight	0.0062			smlhttps	0.0190
	water	0.0052			stripin	0.0175

Note: The table presents the three most likely words in the first four topics generated by the LDA algorithm. The highly technical nature of the documents is obvious here. It is possible to associate overall labels to some of the topics, but not to all. Some topics also appear to contain many filler words (e.g. topic 4 in both patents and articles). Full topics are shown in Appendix A.

(a) Patents

(b) Articles

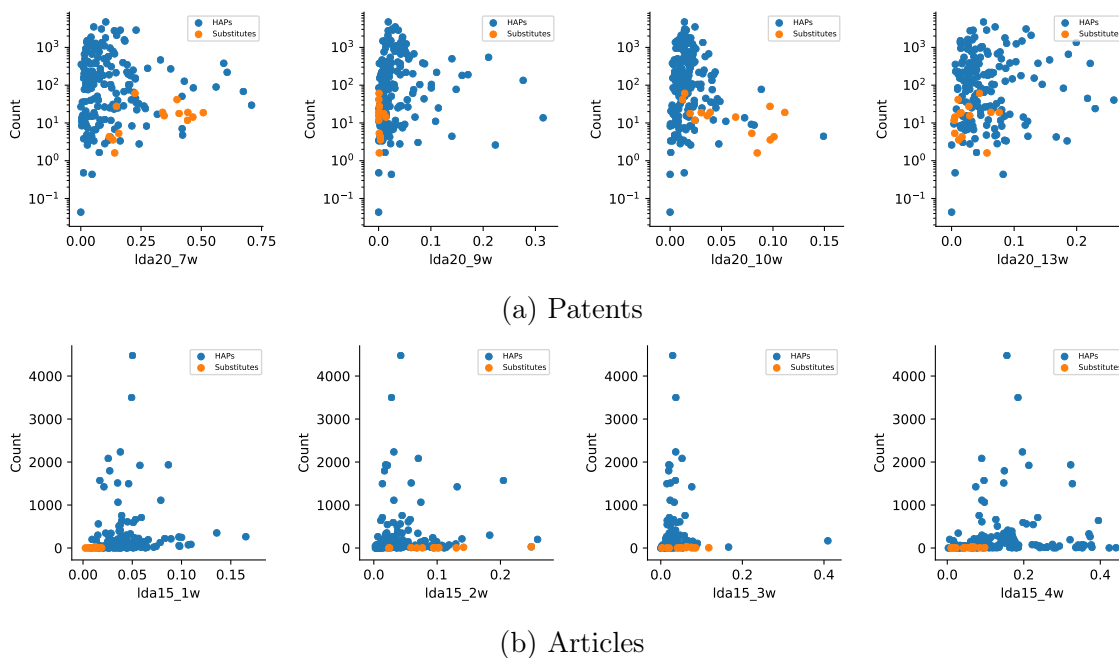Figure 5: Scatterplot of topics proportion and count for selected topics.

Note: The graphs illustrate the usefulness of topic proportions. The scatter plots indicate that there are clear outlier molecules: molecules with semantic contexts far from CFC substitutes. Using topic proportions in the DiD and SCM therefore provides a way of controlling for variation in how different molecules are mentioned in the text of documents.

4 Results

4.1 Difference-in-Differences

I select the most appropriate HAPs to be part of the control group by following two steps. First, I only keep HAPs that have an average pre-period count of the same order of magnitude than the pre-period average count of CFC substitutes (specifically, not higher than 10 times). Second, I rank all remaining HAPs according to how close their pre-trend is to that of the average CFC substitute. I then keep the 28 HAPs with most similar slope. I choose the number of HAPs to be *28* so to construct a control group that is twice as large as the treated group. Figure 6 shows that pre-trends for the treated and control groups appear parallel.
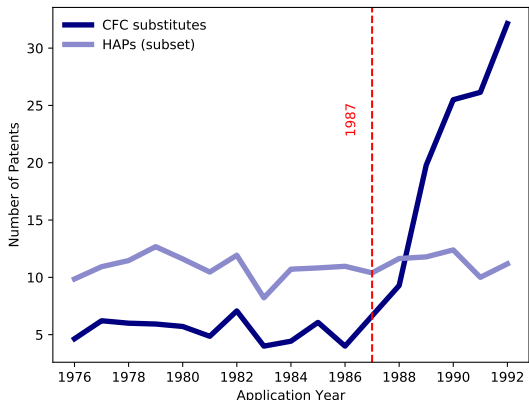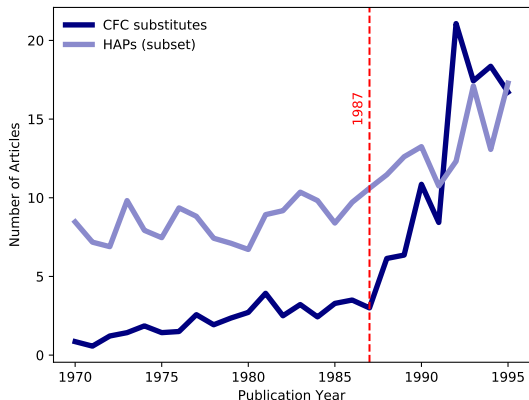
| (a) Patents | (b) Articles |

Figure 6: Pre-trends in counts of documents mentioning CFC substitutes and HAPs

Note: The graphs display the pre-trends for the treated group (CFC substitutes) and the control group constructed using of a subset of the HAP molecules that have counts and pre-trends closest to the average CFC substitutes.

I estimate the following two DiD models: a mean shift specification (Equation 3) and a trend-break specification (Equation 4)

$$Count_{mt} = \alpha + \beta_0 \cdot D_m \cdot Post_t + \lambda_t + \lambda_m + \boldsymbol{\gamma_t} \cdot \boldsymbol{X}_{mt} + \epsilon_{mt} \tag{3}$$

$$Count_{mt} = \alpha + \beta_1 \cdot Y \cdot Post_t \cdot D_m + \beta_2 \cdot Y \cdot Post_t + \beta_3 \cdot Y + \lambda_t + \lambda_m + \boldsymbol{\gamma_t} \cdot \boldsymbol{X}_{mt} + \epsilon_{mt} \tag{4}$$

$Count_{m,t}$ stands for the number of documents mentioning molecule m in year t; $Post_t$ equals one when $t > 1987$; D_m equals one if the molecule belongs to the treated group; Y is a continuous

variable indicating the number of years relative to 1987; λ_m are molecule fixed effects; λ_t are year fixed effects; $\boldsymbol{X}_{mt}$ is a vector of topic proportions. β_0 identifies the DiD estimate. The primary hypothesis is that β_0 and β_1 are positive. Furthermore, Table B1 in Appendix B displays balance tables for topic proportions, and highlights that, for most topics, proportions are significantly different across the two groups. They may help, therefore, capture relevant variation.

The data is particularly suited to a count model. A Zero-Inflated Negative Binomial model is more appropriate than Poisson here because the data is zero-inflated and over-dispersed (as can be seen on Figure B5 in Appendix B). Since results don't differ much from the simpler specification with counts in levels, I report the latter as the main specification for ease of interpretation, and provide results with a Zero-Inflated Negative Binomial model as robustness checks in Section 4.2.

Table 6: Difference-in-differences for CFC substitutes

	(1) Patents	(2) Patents	(3) Patents	(4) Patents	(5) Articles	(6) Articles	(7) Articles	(8) Articles
Post 1987 x Substitutes	16.543***	21.124***	0.455	3.443	7.581***	12.634***	-0.273	1.437
	(1.740)	(2.056)	(2.681)	(2.494)	(1.118)	(1.689)	(1.625)	(2.485)
Post 1987 x Substitutes x Years			5.235***	5.773***			1.825***	2.057***
			(1.005)	(0.825)			(0.379)	(0.500)
Substitutes x Years			0.045	-0.004			-0.027	0.109
			(0.117)	(0.133)			(0.078)	(0.159)
Years			-0.073	-0.086			0.183***	0.212***
			(0.063)	(0.066)			(0.029)	(0.042)
Post 1987			1.182*	1.205			0.968**	0.913*
			(0.647)	(0.760)			(0.449)	(0.515)
Year FE	Yes	Yes	No	No	Yes	Yes	No	No
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Topics (weighted)	No	Yes	No	Yes	No	Yes	No	Yes
Bootstraped	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.841	0.859	0.854	0.875	0.634	0.640	0.634	0.638
Observations	714	595	714	595	1092	846	1092	846

Standard errors in parentheses
Years are relative to 1987.
Time-series: (1976-1992) for patents; (1970-1995) for articles
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 displays the main results. Model 1 corresponds to the differences-in-differences specification without any controls. The binary variable "Post 1987 x Substitutes" equals 1 for observations belonging to the group CFC substitutes after 1987. Model 2 controls for topic proportions and is

the preferred specification. For patents, the coefficient is large and statistically significant: it corresponds to an additional 21 patents on average per year per substitute from 1988 to 1992. Since there are 14 CFC substitutes in my sample, this implies 294 additional patents a year for CFC substitutes in aggregate, and it is equivalent to almost a 400% increase relative to the pre-period mean number of patent (which equals 5.4).

The coefficient for articles is also large and significant: it corresponds to an additional 13 articles on average per year per substitute from 1988 to 1995. Since there are 14 CFC substitutes in my sample, this implies 182 additional patents a year for CFC substitutes in aggregate, and it is equivalent to a 576% increase relative to the pre-period mean number of patent (which equals 2.19). Model 3 and 4 present trend-break specifications and confirm that the data can modeled as a trend break too. The number of patents mentioning CFC substitutes increases with the years after 1987 by 5.8 patents more than the control group. Similarly, the number of articles mentioning CFC substitutes increases with the years after 1987 by 2 articles more than the control group.

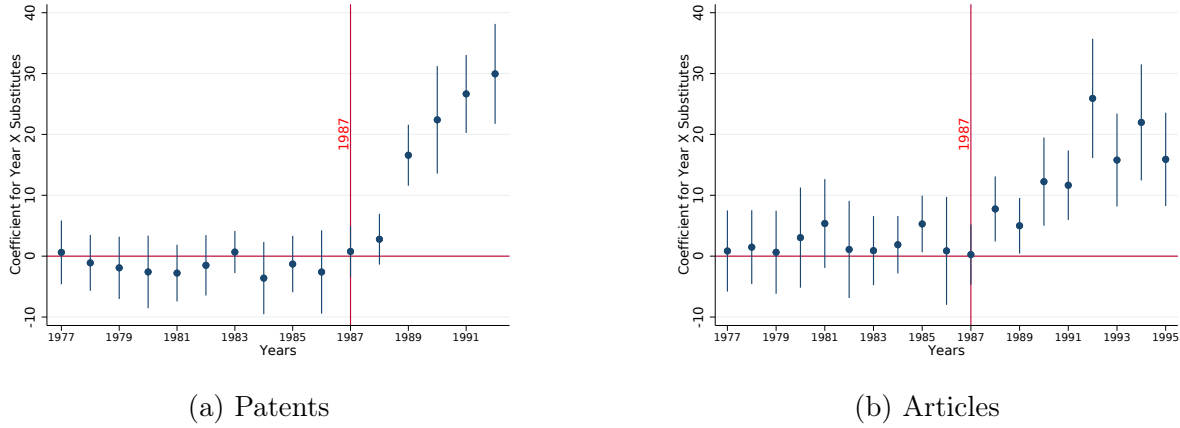(a) Patents   (b) Articles

Figure 7: Difference-in-differences treatment effects by year (controlling for topic proportions)

Figure 7 displays the year-by-year coefficients corresponding to Model 2. We note that, for patents, the treatment effect is statistically significant starting in 1989 and keep increasing with time. We should expect a delay between the moment firms and inventors decide to redirect their efforts towards CFC substitutes and the moment when they are ready to apply for a patent. However, the time required to obtain any technology worth patentable can vary broadly from technology to technology, even within the same technological sector. We can expect some patents to be "low-hanging fruits", i.e. inventions requiring just a few months of R&D work before applying

for a patent. Hence, it may not be surprising to observe a significant treatment effect as early as two years after Montreal. In fact, others have documented similarly fast treatment effects. In the context of energy patenting, Popp (2002) estimates that the mean lag between a change in energy prices and patenting on energy-saving technologies occurs in 3.71 years.

For articles, the treatment effect remains significantly different from zero from 1990 onwards (see Panel b on Figure 7). Since the data only contains the publication date of articles (as opposed to submission date), the lag between treatment and its effect also accounts for processing and reviewing time at academic journals (which, in chemistry and engineering, takes around 9 months on average (Björk et al. 2013)).

4.2   Robustness Checks

I run additional DiD specifications controlling for lags of count. A typical model of science and innovation is one with positive knowledge externalities: patents or articles lead to more patents and articles as scientists and inventors build on previous work. Controlling for lagged count allows me to capture this possible cumulative mechanism. Model 2 in Table 7 confirms that the treatment variable remains statistically significant in both patents and articles. Model 3 shows that results are robust to considering, as an alternative outcome variable, patent and article counts weighted by the number of forward citations each document receives. Citations can be interpreted as the publication's degree of influence (or "quality"). Panel a and c on Figure 8 show that pre-trends in citation weighted counts look similar across CFC substitutes and HAPs for both patents and articles. In a similar fashion, Model 4 shows that the results are robust to considering patent and article counts weighted by the number of times the molecule appeared in each document (molecule "occurrence"). Panel b and d on Figure 8 show that pre-trends in occurrence-weighted counts look similar across CFC substitutes and HAPs in patents. For articles, a small pretrend indicates that, when articles contained the name of CFC substitutes, they tended to mention these substitutes more often over the years. I provide more details about trends in counts for different threshold of occurrences in Figure B4.

Table 7: Robustness checks difference-in-differences

(a) Patents

	(1) Count	(2) Count	(3) Cit	(4) Occ	(5) Cit-Occ
Post 1987 x Substitutes	21.124***	12.999***	29.849***	33.454***	44.910***
	(2.056)	(1.706)	(3.376)	(3.786)	(6.697)
Count (lag 1)		0.385***			
		(0.074)			
Count (lag 2)		0.265***			
		(0.072)			
Year FE	Yes	Yes	Yes	Yes	Yes
Molecule FE	Yes	Yes	Yes	Yes	Yes
Topics (weighted)	Yes	Yes	Yes	Yes	Yes
Bootstraped	Yes	Yes	Yes	Yes	Yes
R-squared	0.859	0.897	0.784	0.702	0.664
Observations	595	528	595	595	595

Time span: 1976 to 1992

(b) Articles

|	(1) Count	(2) Count	(3) Cit	(4) Occ	(5) Cit-Occ
Post 1987 x Substitutes	12.634***	5.098***	12.216***	17.619***	18.110***
	(1.689)	(1.250)	(3.383)	(2.708)	(4.430)
Count (lag 1)		0.341***			
		(0.054)			
Count (lag 2)		0.340***			
		(0.076)			
Year FE	Yes	Yes	Yes	Yes	Yes
Molecule FE	Yes	Yes	Yes	Yes	Yes
Topics (weighted)	Yes	Yes	Yes	Yes	Yes
Bootstraped	Yes	Yes	Yes	Yes	Yes
R-squared	0.640	0.751	0.342	0.490	0.373
Observations	846	790	846	846	846

Time span: 1970 to 1995

Note: The tables present regression results for robustness checks for the difference-in-differences. The specifications use different outcome variables. Column 1 and 2 use counts; column 3 uses citation-weighted counts; column 4 uses occurrences-weighted counts; and column 5 uses counts weighted by both citation and occurrences.
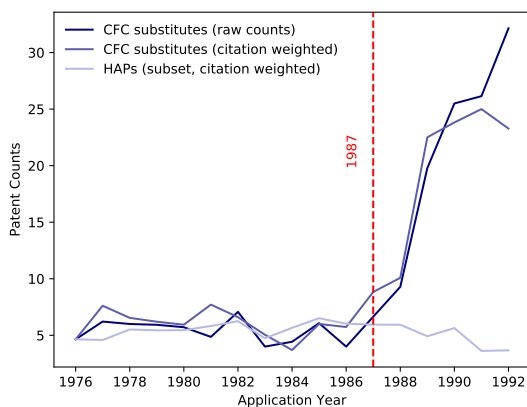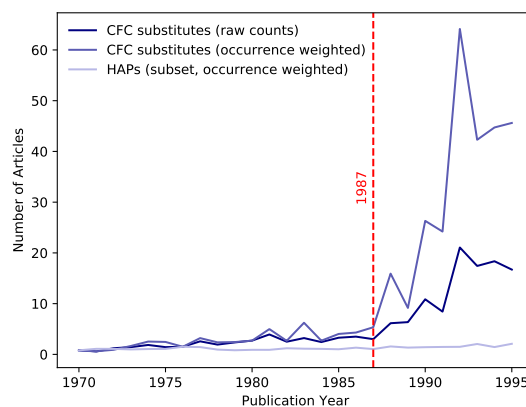
(a) Patent Citation Weighted Counts
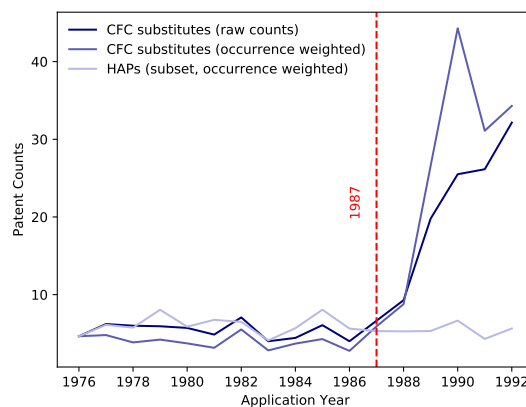
(b) Patent Occurrence Weighted Counts
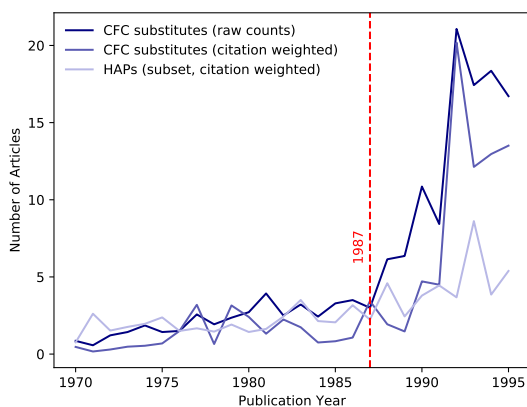
(c) Article Citation Weighted Counts

(d) Article Occurrence Weighted Counts

Figure 8: Time series of citation- and occurrence-weighted counts

Note: The graphs illustrate that weighting counts by the number of citations or the number of times molecules appear in the text does not affect the main uptrend for CFC substitutes relative to HAPs after 1987. Time-series are scaled to make them equal in the first year of the sample.

4.3 Synthetic Control Method

4.3.1 SCM Implementation

The DiD strategy implemented above assumes that the counts of patents and articles for each molecules are independent. However, several CFC substitutes are often mentioned in the same document. For example, in my sample, almost 40% of patents mention more than one molecule[38]. An alternative strategy, here, is to consider adding up the observed counts of the 14 CFC substitutes, and considering them as one single treated unit (what I refer to as the "aggregate CFC

---

[38]Out of 3270 patents mentioning CFC substitutes, 1234 mention more than one CFC substitutes. The DiD considers 5999 observations when, in reality, there are only 3270. For articles, out of 998, 226 mention more than one CFC substitutes. Hence the DiD considers 1266 observations when, in reality, there are only 998.

26

substitute"). To illustrate, Figure 9 plots the number of patents mentioning each CFC substitute, and the thick line corresponds to the number of patents mentioning any of the 14 CFC substitutes.

To obtain an estimate of the treatment effect on this aggregate quantity, I implement a Synthetic Control Method and constructs a counterfactual molecule that mimics the evolution of the CFC substitutes in aggregate. SCM was specifically developed to evaluate the effects of large aggregate interventions when the treatment affects aggregate quantities (Abadie et al. 2003, 2010, 2015; Athey et al. 2017). Many interventions are in fact implemented at an aggregate level and have an impact on a small number of large entities, such as cities, school districts, or states. I enlarge the application of SCM to a new kind of aggregate entity: field of scientific and engineering inquiry.
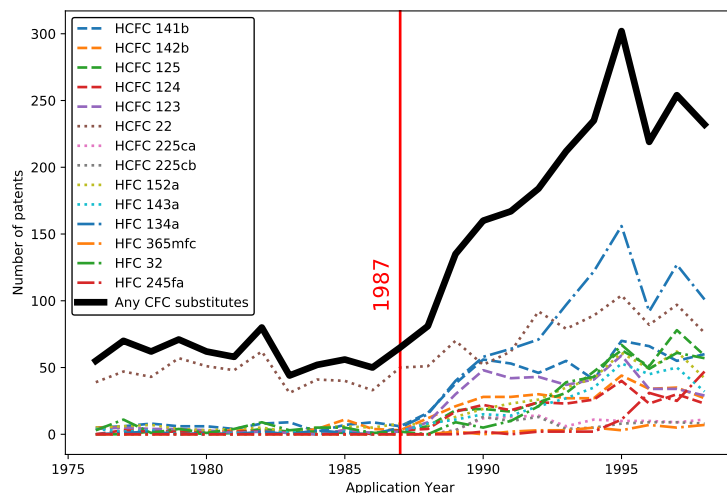

Figure 9: Patent counts for CFC substitute, individually and aggregated

Note: The graph illustrates the difference between considering the 14 molecules independently and considering them as one treated molecule. Since the names of different CFC substitutes often appear simultaneously in the same documents, the individual time series of each CFC substitute are not independant from each other. The thick line called "Any CFC substitutes" corresponds to the number of patents mentioning any of the 14 CFC substitutes. It is equivalent to considering the 14 compounds as one and only one molecule. I implement the synthetic control method on this "aggregated CFC substitute". Figure C1 displays a similar graph for articles.

The SCM consists in using a weighted average of a set of control units with the weights chosen so that the weighted average is similar to the treated unit regarding covariates and outcome in the pretreatment periods[39]. The outcome, here, is the number of patents (or articles) that mention any of the 14 molecules. The synthetic control is constructed by fitting the values of counts in

---

[39]Appendix C provides more details on the theoretical foundations for the SCM.

the pretreatment period and the topic proportions. The treatment year is the first year in which the treatment becomes active: this is defined as 1988 since Montreal was agreed in 1987. To be conservative, I use data only until 1985 to fit the synthetic control. Topic proportions are averaged over the entire pre-1985 period, while the outcome, count, is not.

As explained by Abadie et al. (2015), reducing the size of the donor pool can limit the risk of over-fitting as well as the risk of interpolation biases. Following their advice, I use a smaller donor pool containing only the thirty HAPs closest to the treated unit in terms of counts[40]. For inference, I follow the non-parametric approach method suggested by Abadie et al. (2010) and Abadie et al. (2015). The approach is akin to implementing placebo tests wherein each unit in the control group is assumed to have received the treatment at the year 1987. The "true" treatment effect can then be compared to the distribution of placebo treatment effects. A p-value is calculated as the fraction of placebo effects that are greater than or equal to the effect estimated for the "true" treated unit. As suggested by Abadie et al. (2010), I compute the ratios of post-RMSPE over pre-RMSPE and examine where in the distribution of those ratios, the treated unit lies[41].

4.3.2 SCM Results

Figure C4 graphically displays the results of the SCM for CFC substitutes. The graphs on the left-hand side represent the raw effect, that is the observed time series of the treated group along with the time series of the constructed control. On the right-hand sides are shown the placebo tests to evaluate the significance of the results; the black lines show the effect on the treated group relative to the control group, while each gray line is a placebo test performed on a unit drawn from the donor pool. The treatment effect on CFC substitutes appears significant for both patents and articles. We note that the black line rises above most other lines mostly as from 1989. This indicates that, similarly as in the DiD, the treatment effect is statistically significant starting in 1989. For articles, the treatment effect is statistically significant starting in 1992.

---

[40]I also check that there is no risk of extrapolation. See Table C1 in Appendix C

[41]The pre-RMSPE measures lack of fit between the path of the outcome variable for any particular unit and its synthetic counterpart: the pre-RMSPE of unit 1 is defined as $(\frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}))^{1/2}$ where T_0 is the number of pretreatment periods. A post-RMSPE can be similarly defined for periods going from $T_0 + 1$ to the end of time-series available.

The overall treatment effects (graphically, the area between the two curves on the left-hand side graphs) correspond to 84 (resp. 49) additional patents (resp. articles) per year on average between 1988 and 1992 (resp. 1995). This roughly corresponds to 135% and 177% increases in patents and articles. To examine more closely how the synthetic unit is composed, Figure 8 provides information regarding the main HAPs entering the synthetic control[42]. We note that the SCM picked up HAPs that have very broad industrial applications (not unlike CFC substitutes). In particular, we find similar applications (e.g. coatings or solvents).

(a) Patents: raw effect (left) and placebo tests (right)

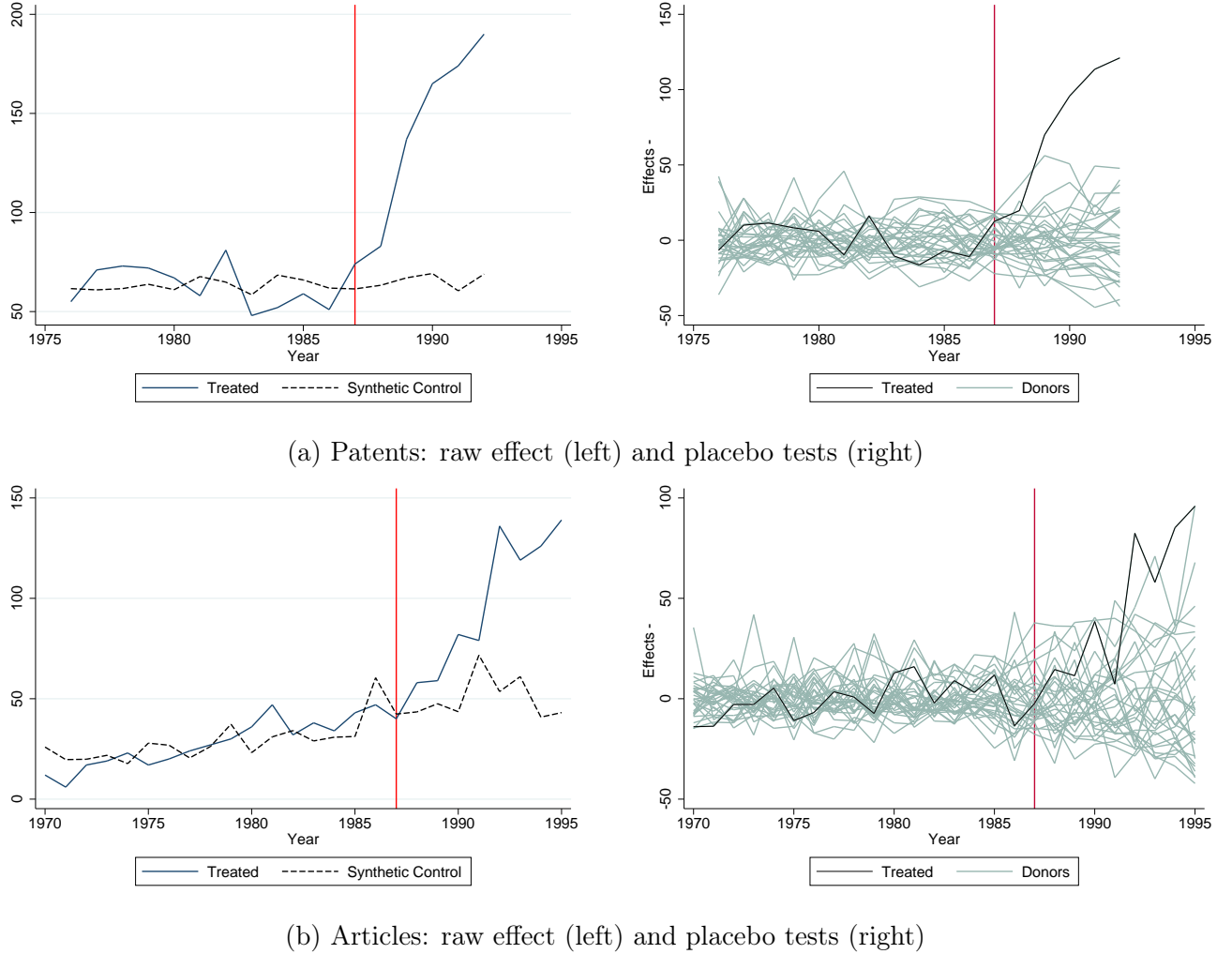(b) Articles: raw effect (left) and placebo tests (right)

Figure 10: SCM graphs for CFC substitutes

---

[42]Additionaly, Table C2 displays the value of each variable's contribution to the synthetic control.

Table 8: HAPs contributing to the synthetic control

(a) Patents

HAPs	Weight	Description
Calcium cyanamide	0.327	Used as a fertilizer, defoliant, herbicide, fungicide, and pesticide; in the manufacture and refining of iron; and in the manufacture of calcium cyanide, melamine, and dicyandiamide.
Polychlorinated biphenyls	0.206	Group of chemicals characterized by non-flammability, stability, high boiling point and electrical insulating properties. Hundreds industrial applications: electrical and heat transfer, paints, plastics.
Methyl bromide	0.140	Used as a fumigant in soil to control fungi, nematodes, and weeds; inspace fumigation of food commodities (e.g., grains); and in storage facilities (such as mills, warehouses, vaults, ships, and freight cars) to control insects and rodents.
Benzidine	0.116	Production of dyes, especially azo dyes in the leather, textile, and paper industries
o-Xylenes	0.103	Used in the production of ethylbenzene, as solvents in products such as paints andcoatings, and are blended into gasoline.

(b) Articles

HAPs	Weight	Description
Bromoform	0.503	Used as a fluid for mineral ore separation, as a laboratory reagent and in the electronics industry in quality assurance programs. Was used as a solvent for waxes, greases, and oils, as an ingredient in fire-resistant chemicals and in fluid gauges. Also used as an intermediate in chemical synthesis, as a sedative and cough suppression agent.
1,4-Dichlorobenzene	0.332	Used mainly as a fumigant for the control of moths, molds and mildews, and as a space deodorant for toilets and refuse containers. Also used as an intermediate in the production of other chemicals, in the control of tree-boring insects, and in the control of mold in tobacco seeds.
Trifluralin	0.165	Herbicide. Mostly used on cotton, soybeans and some fruits and vegetables

Note: The tables describe the HAPs entering the synthetic control for the SCM specification. The information displayed in the "Description" column was collected from the EPA website.

5 Mechanisms and Alternative Hypotheses

5.1 Mechanisms

In this section, I provide more descriptive details about the increase in research and innovation fostered by Montreal.

In the Appendix, Figure D1 that the increase in the number of patents mentioning CFCs applies to all countries; we note a particularly strong increase for patents with assignees located in Japan and the UK. Table D2 shows that, both before and after Montreal, US-based private sector firms are the entitites patenting the great majority of CFC substitutes. Finally, Figure D2 illustrates that the patent codes most frequent before 1987 tend to also be the most frequent after 1987. At the same time, many codes with low or no frequency before 1987 become important after 1987[43].

I also document that the post-Montreal burst of innovations on CFC substitutes is not driven by a few firms that would have been historically patenting on CFC substitutes since the 1970s. Figure 11a displays the yearly number of unique assignees with patents mentioning CFC substitutes and HAPs. It indicates that the post-Montreal patenting world features many more firms. Figure 11 displays the yearly number of assignees that are "new", meaning they appear for the first time in the data with a patent mentioning CFC substitutes and HAPs. The figure confirms that, after 1987, many firms with no prior experience on CFC substitutes begin patenting.

5.2 Were CFC Substitutes Kept "Secret"?

In this section, I further examine whether firms initiated the transition to CFC substitutes before the signature of the Montreal Protocol, without patenting but instead keeping their technologies as trade secrets. At the end of the 1970s, a few firms announced some R&D investments into CFC substitutes. Although the same firms, soon after, announced terminating those same R&D

---

[43]C08G: Macromolecular compounds obtained otherwise than by reactions only involving carbon-to-carbon unsaturated bonds, C10M: Lubricating compositions; Use of chemical substances either alone or as lubricating ingredients in a lubricating composition, C23G: Cleaning or de-greasing of metallic material by chemical methods other than electrolysis, C11D: Detergent compositions; Use of single substances as detergents; Soap or soap-making; Resin soaps; Recovery of glycerol.
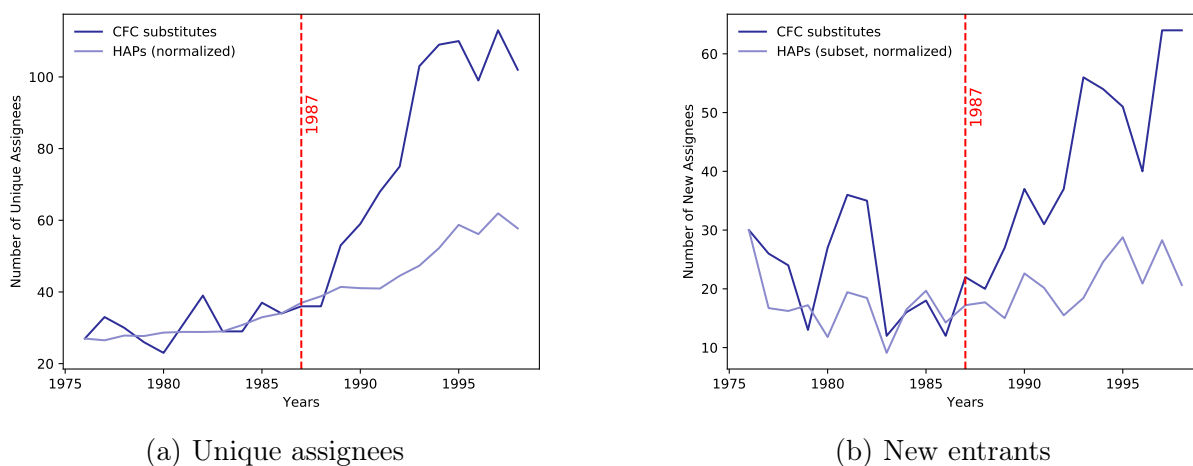
(a) Unique assignees (b) New entrants

Figure 11: Number of unique and "new entrant" patent assignees

programs, it has been suggested that they developed key technologies which they kept secret.

Here, I argue that, if that had been the case, we should expect a one-time increase in patent counts in the immediate aftermaths of Montreal. Indeed, when competitors work on closely related projects, delaying patenting sharply increases the risk that competition patents first. If, until Montreal, firms were with the impression that policy pressure was low, they may have elected not to patent. However, once the protocol is signed, they have clear incentives to patent any old technologies that they may have previously developed as fast as possible to outrun possible competitors.

Figure 12 plots the number of patents mentioning CFC substitutes month by month in the two years that followed Montreal. On the first graph, we note the absence of a patenting peak after 1987. On the second graph, I present trends for assignees that never patented on CFC substitutes before 1987 and those who did. If the R&D carried out before Montreal was a key driver to the post-Montreal increase in patenting, we would observe major differences in the patenting trends of firms with prior patenting experience on CFC substitutes and those without. Although a gap seems to build up over time, trends look mostly similar.

Figure 13 illustrates the possible key role of a few manufacturers: the scatter plot shows,

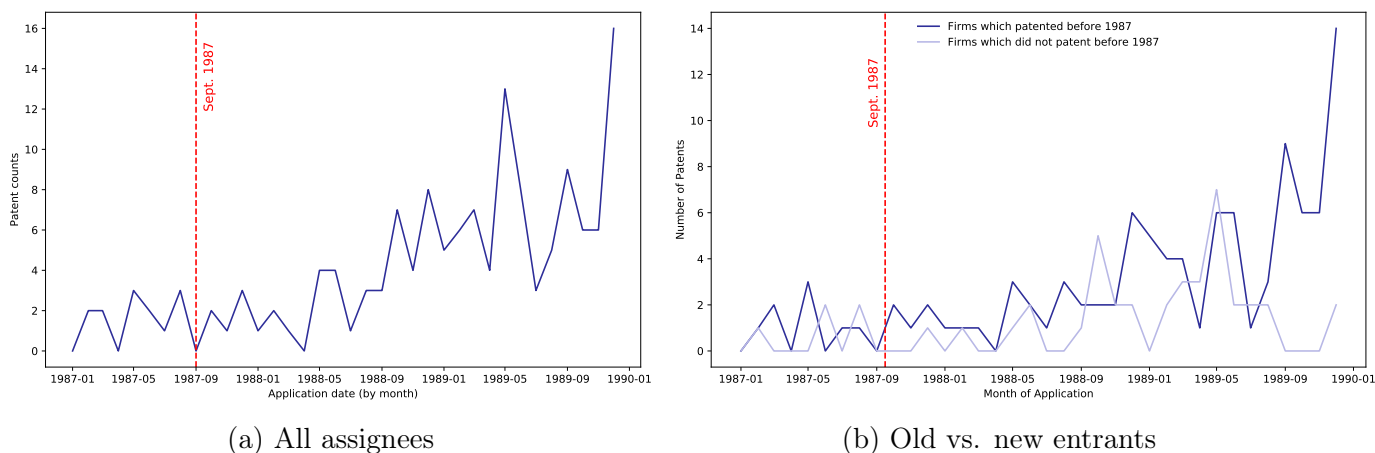(a) All assignees   (b) Old vs. new entrants

Figure 12: Monthly counts for patents mentioning CFC substitutes

Note: The graphs show the monthly trends in count of patents mentioning CFC substitutes. Panel 12b shows the monthly count of patents mentioning CFC substitutes for firms which patented on CFC substitutes before 1987 vs. those who did not. The year used is the application year. The period "Before 1987" includes the year 1987. To limit noise, the sample used to generate the table contains only documents with at least three occurrences of CFC substitutes.

for each firm in the sample, the number of patents between 1975 and 1986 on the x-axis and the number of patents in the two years that followed Montreal on the y-axis. We see that a positive trend is mostly driven by two firms: DuPont and Allied. Excluding those, there are no clear correlations between patenting before 1987 and patenting in the immediate aftermaths of Montreal. This plot, however, motivates a more detailed investigation in the behavior of DuPont and Allied.

Figure 14a shows that most patents granted to DuPont and Allied were applied for after 1989, and in particular, Figure 14b shows that there is no sudden peak patenting right after Montreal. Instead, we observe a gradual ramping up of patenting activity. Figure 14c illustrates that the patents granted to DuPont and Allied which received the greatest number of citations mostly originate from 1989 to 1991. Figure 14d indicates, however, that, in the weeks that followed Montreal, both DuPont and Allied applied for patents that would go on receiving a high number of citations. This seems to indicate that DuPont and Allied likely had a first mover advantage on some technologies. However, the magnitude of the ramping up in patenting activity that follows from 1990 onwards allows concluding that most of the innovative activity started after Montreal[44].

---

[44]Another way of examining the effect of the international agreement on DuPont would be to look at DuPont's stock market valuation. Unfortunately, although in 1986 DuPont produced CFCs for about half of the US market, it represented only 2.2% of DuPont revenues (1.8% in 1984 and 1.7% in 1985), 2% of corporate assets and 0.9%
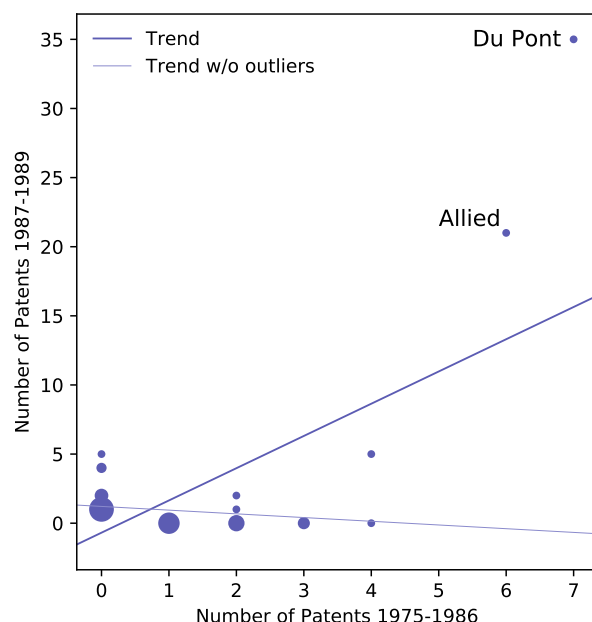
Figure 13: Scatterplot of patenting activity before and after 1987

Note: The size of the dot is proportional to the number of firms. The year used is the application year. To limit noise, the sample used to generate the table contains only documents with at least three occurrences of CFC substitutes.

5.3 Was Consumer Pressure a Potential Driver?

The science of ozone made much progress during the 1980s: in 1985, scientists detected a large depletion of ozone over Antarctica (the "hole"), and importantly, were able to causally attributed it to CFCs in March 1988 (the "discovery"). The image of the Earth seen from space with a massive hole (artifically colored in blue for the occasion) became world-famous and moved public opinion. The perceived benefits of phasing-out CFCs certainly increased, and made the issue more salient in the eyes of public opinion and consumers. To what extent, then, did consumer pressure drive innovation in the aftermaths of Montreal?

There exists only few empirical analysis suggesting that consumer pressure can be effective (Lyon et al. 1999; Popp et al. 2011), and they all deal with local pollutants such as toxic chemical emissions. We could reasonably expect consumer pressure to be less effective for a global air pollutant such as ozone. As profit-maximizing entities, firms would have few incentives to incur

of DuPont's employees (Reinhardt et al. 1989). It is therefore unlikely that financial markets would capture much impact. Additionaly, it would difficult to attribute any movement to the regulation of CFCs only and not to other parts of DuPont's business (especially since DuPont was facing other public relations issues related to medical implants of which it supplied the raw material).

(a) Yearly Counts

(b) Monthly Counts

(c) Yearly Citation Weighted Counts

(d) Monthly Citation Weighted Counts
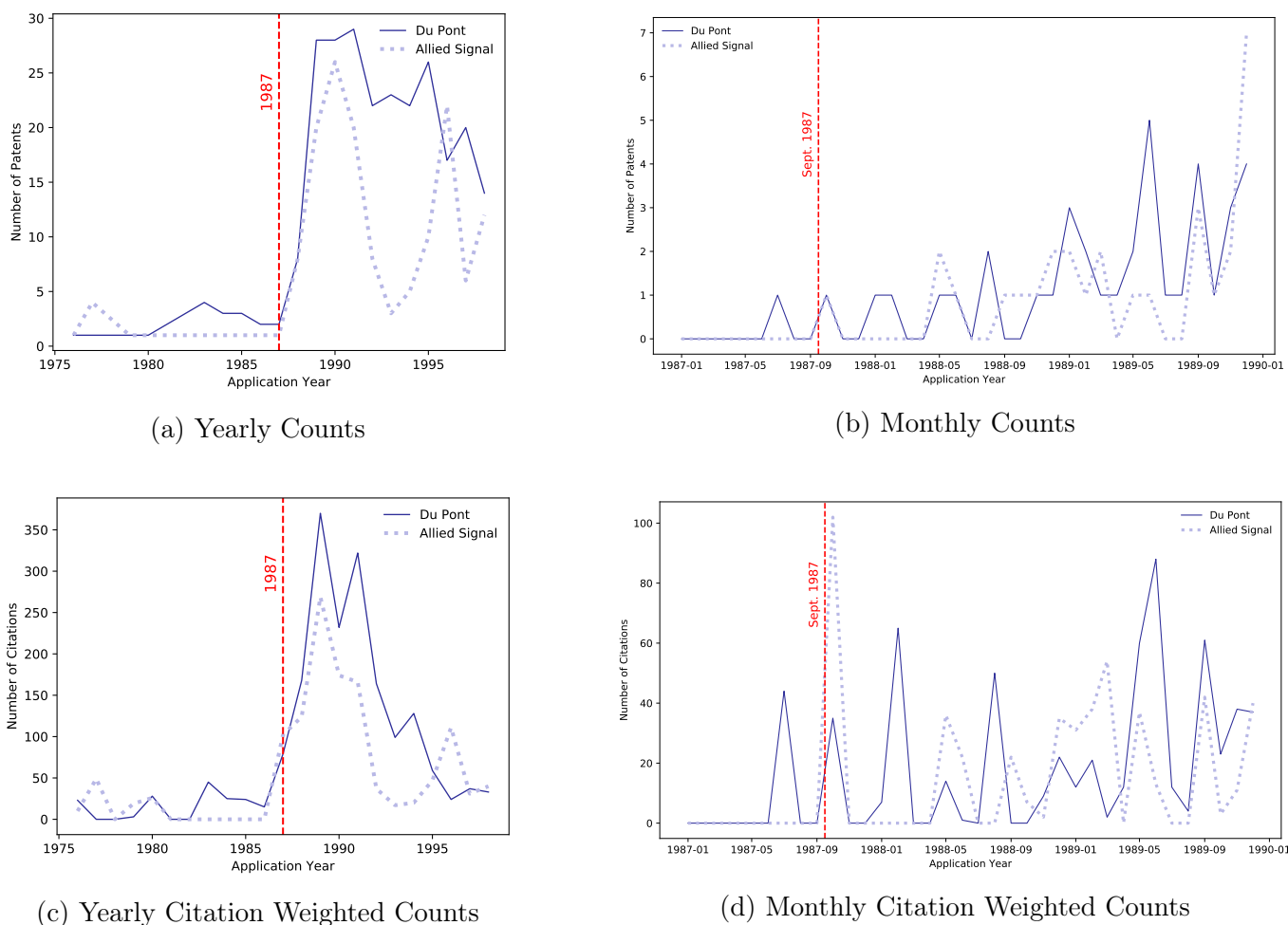
Figure 14: Patent counts for DuPont and Allied

Note: Figure 14a shows that most patents granted to DuPont and Allied were applied for after 1989. Figure 14b shows that there is no sudden peak patenting right after Montreal. Instead we observe a gradual ramping up of patenting activity. Figure 14c illustrates that the patents granted to DuPont and Allied which received the greatest number of citations mostly originate from 1989 to 1991. Figure 14d indicates, however, that, in the weeks that followed Montreal, both DuPont and Allied applied for patents that would go on receiving a high number of citations. Only patents with at least 3 occurrences of a molecule are retained in the sample.

R&D costs without the guarantee of a large market, and without the guarantee that their foreign and domestic competitors do the same. To investigate the role of consumer pressure for innovation on CFC substitutes, I leverage the fact that not all molecules or applications were exposed to consumers. Indeed, examining the effect of Montreal on patenting related to substitutes unexposed to consumers would allow to set aside the possible influence of public opinion and consumer pressure.

I use data available on the EPA SNAPs website to identify which CFC substitutes in my sample were not exposed to consumers. The EPA website lists a total of 1001 requests corresponding to

a given substance for a given application[45]. I manually classify applications as either consumer exposed, not consumer exposed or undetermined. Doing so, I identify 6 CFC substitutes that are credibly not consumer exposed. The typical applications requested for these CFC substitutes are all related to industrial activities invisible to consumers: e.g., centrifugal chillers, foam blowing agents for rigid polyurethane, or precision cleaning.

Figure 15 displays the time-series of the number of patents mentioning CFC substitutes that were arguable not exposed to consumers. We observe that, for almost all of them, the number of patents increases strongly after 1987. This is indicative that consumer pressure and public opinion did not play the major role in driving innovation in the aftermath of the Montreal protocol.
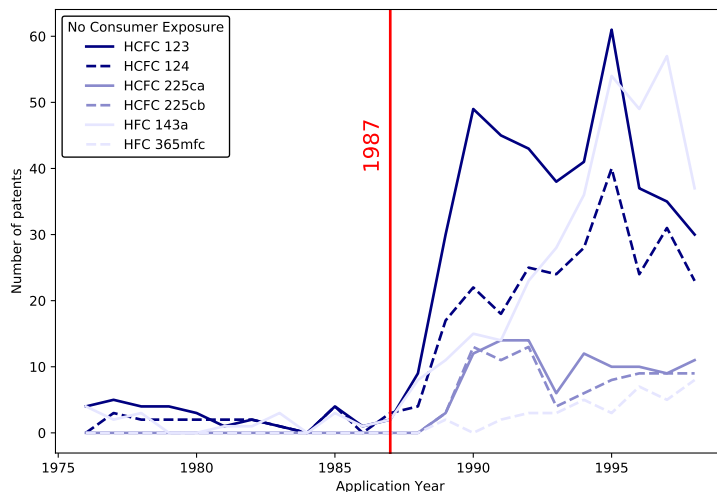
Figure 15: Strong Increases for Molecules Not Exposed to Consumers

6   Discussion and Concluding Remarks

Tackling environmental problems often relies on developing and diffusing new technologies. It is, therefore, important to better understand the drivers of technological change. In this paper, I document that the Montreal Protocol, and its following amendments, led to the development of CFCs substitutes. This empirical evidence goes against the often-heard narrative that alternatives technologies were readily available before the treaty. In fact, that narrative was once characterized

---

[45]As part of the implementation of the Montreal protocol, the EPA was in charge of controlling which substitutes firms could use. To this end, any firm using or producing substitutes had to request an authorization for a given substitute in a given application. An example is HCFC-225cb for Electronics Cleaning.

by sociologist Reiner Grundmann as "the most pervasive and most widespread myth surrounding the Montreal Protocol" (Grundmann 1998).[46]. And indeed, the treatment effect that I estimate tells a story where almost all of the science and innovation on CFC substitutes was triggered by the post-Montreal regime. The magnitude of the effect is even consistent with what has been described as a "burst of industrial creativity" (Meadows et al. 1992).

As explained, CFC substitutes did exist on paper for a long time, but the challenge was to make them cost effective for several industrial applications. In this context, patenting activity can be interpreted as efforts to lower down such costs, the typical reason why firms patent in response to environmental policies. The signature of Montreal induced firms and researchers to orient their R&D effort towards CFC substitutes. It had the immediate effect of modifying expectations about future prices and created a worldwide demand for substitutes. As a result, it incentivized profit-seeking firms to bring CFC substitutes to market at the cheapest possible cost. Researchers publishing in peer-reviewed journals also redirected their work towards CFC substitutes likely either due to a shift in their personal research priorities in a context where both government and industry were in need of more information about these molecules.

The role of the Montreal Protocol in solving the crisis has been intensely discussed. In particular, in a seminal article, Barrett (1994) developed a theory of international environmental agreements which highlighted that Montreal was sucessfully negotiated because it was cheap to do so. Murdoch et al. (1997) also argued that some countries, and in particular the USA, accepted to sign Montreal because the costs of doing so were low. They highlighted the existence of a cost-and-benefit analysis issued by the EPA in 1987 concluding that the estimated benefits (cancers avoided) overwhelmed the estimated costs on the industry.

The idea that the Montreal agreement was "cheap" has often been rephrased as "substitutes

---

[46]He traced the origin of the "myth" to the fierce opposition between Americans and Europeans during the Concorde controversy. In the early 1970s, France and the UK hoped to conquer the world with supersonic jets, but in 1974, Americans denied the authorization to land the aircraft on U.S. soil. They argued that the pollution emitted by the plane's engine was a serious threat to the ozone layer. The claims were eventually dismissed as it was shown that the atmospheric reactions did not occur at the Concorde's flying altitudes (Benedick 2009, p. 33). But the incident certainly left an aftertaste of distrust for Europeans, and maybe primed them to assume that, when dealing with Americans, environmental issues are but a disguise for commercial interests. CFC regulations, then, came onto the international agenda, and Americans adopted a pro-regulatory stance. For European manufacturers, it was a small step, possibly, to assume secret substitutes existed and that US manufacturers were up to capturing larger market shares.

existed". But it is more exact to state the idea, not in terms of whether substitutes existed or not, but in terms of what, at the times, industries and governments expected to be feasible, technologically speaking. A crucial but typically understated fact, however, is that the targets in Montreal were not ambitious at all. Hence, we shall not be very surprised that the expected compliance costs calculated at the time by the EPA were low. Indeed, the agreement was not successful in its attempt to negotiate a full ban (which the industry actively lobbied against). In terms of emission reductions, it therefore settled for modest targets.

The real success, therefore, should be seen in the later amendments: London in 1990 and Copenhagen in 1992. More ambitious reduction targets were then agreed (deepening), while other molecules were added to the list of regulated compounds (widening). An interesting question to ask, therefore, is how come countries manage to agree to widen and deepen the protocol. To this, my paper proposes a partial answer. Yes, on the one hand, Montreal was not ambitious. But, on the other, it set up a credible enforcement mechanism which was critical to change firms' expectations about CFCs. Hence, the agreement induced innovation. Doing so, it contributed to lowering down the perceived costs associated to a full phase-out, while at the same time, the perceived benefits of phasing-out CFCs certainly increased due to the discovery of the ozone hole.

Hence, the success story of Montreal is better summarized as a repeated cooperation game with ratcheting up ambitions, and where innovation plays the critical role of enabling factor. It is a case of the theory of international environmental agreements and the theory of induced innovation walking hand in hand. At each stage, binding reductions force firms to innovate and develop technologies to comply. But with realised innovation, expectations for further innovation increase as well. The costs of deeper reduction then seem more affordable, and governments and industries become more willing to bind themselves to it.

When solutions to environmental problems are plagued with technological uncertainties or high price tags, decision-makers are incentivized to adopt a "wait-and-see" strategy: wait for proven new technologies, then negotiate an agreement. By showing that agreements can encourage the development of green technologies, this paper suggests they should be negotiated as early as possible if we hope to solve global environmental problems.

7 References

Abadie, Alberto and Javier Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country." In: *The American economic review* 93.1, pp. 113–132.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." In: *Journal of the American Statistical Association* 105.490, pp. 493–505.

— (2015). "Comparative Politics and the Synthetic Control Method." In: *American journal of political science* 59.2, pp. 495–510.

Acemoglu, Daron (1998). "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." In: *The quarterly journal of economics* 113.4, pp. 1055–1089.

Aghion, Philippe et al. (2016). "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry." In: *The journal of political economy* 124.1, pp. 1–51.

Aichele, Rahel and Gabriel Felbermayr (2011). "Kyoto and the Carbon Footprint of Nations." In: *Journal of environmental economics and management* 63.3, pp. 336–354.

Ambec, Stefan et al. (2013). "The Porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness?" In: *Review of Environmental Economics and Policy* 7.1, p. 2.

Athey, Susan and Guido W Imbens (2017). "The State of Applied Econometrics: Causality and Policy Evaluation." In: *The journal of economic perspectives: a journal of the American Economic Association* 31.2, pp. 3–32.

Azoulay, P, J S Graff Zivin, and J Wang (2010). "Superstar extinction." In: *The Quarterly Journal of Economics.*

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin (2019). "Does science advance one funeral at a time?" In: *The American economic review.*

Barrett, Scott (1994). "Self-enforcing International Environmental Agreements." In: *Oxford economic papers*, pp. 878–894.

— (1999). "Montreal Versus Kyoto: International Cooperation and the Global Environment." In: *Global public goods: international cooperation in the 21st century.* Oxford University Press.

— (2003). *Environment and Statecraft: The Strategy of Environmental Treaty-making: The Strategy of Environmental Treaty-making.* Oxford University Press.

Benedick, Richard Elliot (2009). *Ozone Diplomacy: New Directions in Safeguarding the Planet.* Harvard University Press.

Björk, Bo-Christer and David Solomon (2013). "The publishing delay in scholarly peer-reviewed journals." In: *Journal of informetrics* 7.4, pp. 914–923.

Blei, David M (2012). "Probabilistic Topic Models." In: *Communications of the ACM* 55.4, pp. 77–84.

Blei, David M and John D Lafferty (2009). "Topic Models." In: *Text Mining: Classification, Clustering, and Applications* 10.71, p. 34.

Calel, Raphael and Antoine Dechezleprêtre (2016). "Environmental Policy and Directed Technological Change: Evidence from the European Carbon Market." In: *The review of economics and statistics* 98.1, pp. 173–191.

Cohen, Wesley M, Richard R Nelson, and John P Walsh (2000). "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)."

Dekker, Thijs et al. (2012). "Inciting Protocols." In: *Journal of environmental economics and management* 64.1, pp. 45–67.

Finus, Michael and Sigve Tjøtta (2003). "The Oslo Protocol on Sulfur Reduction: The Great Leap Forward?" In: *Journal of public economics* 87.9, pp. 2031–2048.

Grundmann, Reiner (1998). "The Strange Success of the Montreal Protocol-why Reductionist Accounts Fail." In: *International Environmental Affairs* 10.3, pp. 197–220.

Hall, Bronwyn H and Adam B Jaffe (2012). "Measuring Science, Technology, and Innovation: A Review." In: *Report Prepared for the Panel on Developing Science, Technology, and Innovation Indicators for the Future.*

Heal, Geoffrey (2016). *Endangered Economies: How the Neglect of Nature Threatens Our Prosperity.* Ed. by Columbia University Press. Columbia University Press.

Henderson, Rebecca, Adam B Jaffe, and Manuel Trajtenberg (1998). "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988." In: *The review of economics and statistics* 80.1, pp. 119–127.

Hicks, John R (1932). "The Theory of Wages." In: *Et Seq*, p. 54.

Iaria, A, C Schwarz, and F Waldinger (2018). "Frontier knowledge and scientific production: Evidence from the collapse of international science." In: *The Quarterly Journal of.*

Jaffe, Adam B and Karen Palmer (1997). "Environmental Regulation and Innovation: A Panel Data Study." In: *The review of economics and statistics* 79.4, pp. 610–619.

Jaffe, Adam B, Richard G Newell, and Robert N Stavins (2002). "Environmental Policy and Technological Change." In: *Environmental & Resource Economics* 22.1, pp. 41–70.

Kellenberg, Derek and Arik Levinson (2014). "Waste of Effort? International Environmental Agreements." In: *Journal of the Association of Environmental and Resource Economists* 1.1/2, pp. 135–169.

Lyon, Thomas P and John W Maxwell (1999). "'Voluntary' Approaches to Environmental Regulation: A Survey." In: *Ssrn Electronic Journal*.

Meadows, Donella H et al. (1992). *Beyond the Limits: Global Collapse or a Sustainable Future*. Earthscan Publications Ltd.

Miller, Alan S and Irving M Mintzer (1986). "The Sky is the Limit." In: *Washington, DC: World Resources Institute*.

Milliman, Scott R and Raymond Prince (1989). "Firm incentives to promote technological change in pollution control." In: *Journal of environmental economics and management* 17.3, pp. 247–265.

Molina, Mario J and F Sherwood Rowland (1974). "Stratospheric Sink for Chlorofluoromethanes: Chlorine Atom-catalysed Destruction of Ozone." In: *Nature* 249.28, pp. 810–812.

Moser, Petra (2012). "Innovation without patents: Evidence from World's Fairs." In: *The Journal of Law and Economics* 55.1, pp. 43–74.

Murdoch, James C and Todd Sandler (1997). "The Voluntary Provision of a Pure Public Good: The Case of Reduced CFC Emissions and the Montreal Protocol." In: *Journal of public economics* 63.3, pp. 331–349.

— (2009). "The Voluntary Provision of a Pure Public Good and the Montreal Protocol: Behavioral and Data Concerns." In: *Oxford economic papers* 61.1, pp. 197–200.

Parson, Edward A (2003). *Protecting the Ozone Layer: Science and Strategy*. Oxford University Press.

Popp, David (2002). "Induced Innovation and Energy Prices." In: *The American economic review*, pp. 160–180.

— (2005). "Lessons from Patents: Using Patents to Measure Technological Change in Environmental Models." In: *Ecological economics: the journal of the International Society for Ecological Economics* 54.2, pp. 209–226.

— (2019). "Environmental Policy and Innovation: A Decade of Research."

Popp, David, Richard G Newell, and Adam B Jaffe (2010). "Chapter 21 - Energy, the Environment, and Technological Change." In: *Handbook of the Economics of Innovation*. Ed. by Bronwyn H Hall and Nathan Rosenberg. Vol. 2. North-Holland, pp. 873–937.

Popp, David, Tamara Hafner, and Nick Johnstone (2011). "Environmental policy vs. public pressure: Innovation and diffusion of alternative bleaching technologies in the pulp industry." In: *Research policy* 40.9, pp. 1253–1268.

Porter, Michael E and Claas Van Der Linde (1995). "Toward a New Conception of the Environment-competitiveness Relationship." In: *The journal of economic perspectives: a journal of the American Economic Association*, pp. 97–118.

Redner, Sidney (2004). "Citation Statistics from More Than a Century of Physical Review." In: *Arxiv Preprint Physics/0407137*.

Reinhardt, Forest and Richard H K Vietor (1989). "Du Pont Freon Products Division (b)." In: *Managing Environmental Issues: A Casebook*, pp. 261–286.

Röder, Michael, Andreas Both, and Alexander Hinneburg (2015). "Exploring the Space of Topic Coherence Measures." In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. WSDM '15. Shanghai, China: ACM, pp. 399–408.

Sampat, Bhaven N (2018). *A Survey of Empirical Evidence on Patents and Innovation*. Tech. rep. w25383. National Bureau of Economic Research.

Sunstein, Cass R (2007). "Of Montreal and Kyoto: A Tale of Two Protocols." In: *Harv. Envtl. L. Rev.* 31, p. 1.

Thompson, Peter and Melanie Fox-Kean (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." In: *The American economic review*, pp. 450–460.

Wagner, Ulrich J (2009). "The Voluntary Provision of a Pure Public Good? Another Look at Cfc Emissions and the Montreal Protocol." In: *Oxford economic papers* 61.1, pp. 183–196.

— (2016). "Estimating Strategic Models of International Treaty Formation." In: *The Review of economic studies* 83.4, pp. 1741–1778.

Wang, Dashun, Chaoming Song, and Albert-László Barabási (2013). "Quantifying Long-term Scientific Impact." In: *Science* 342.6154, pp. 127–132.

Williams, Heidi L (2013). "Intellectual Property Rights and Innovation: Evidence from the Human Genome." In: *The journal of political economy* 121.1, pp. 1–27.

— (2017). "How do patents affect research investments?" en. In: *Annual review of economics* 9, pp. 441–469.

A   Appendix A: Cleaning Procedures and Topic Modelling

A1   Cleaning procedure

A1.1   PatentS

- Cleaning steps to search and count the number of times a molecule name appear in the text:

    – Lowercase
    – Replace the following punctuation signs by an empty string: , - ( )
      For example, '3-Amino-2,5-dichlorobenzoic acid' becomes '3amino25dichlorobenzoic acid'
    – Replace any other type of punctuation by a space

- Cleaning steps to transform the text into a list of words (necessary for topic modeling)

    – Normalize hyphenated words
    – Normalize quotation marks
    – Normalize unicode strings
    – Replace any punctuation by a space
    – Lowercase
    – Replace any number by the string '_NUMBER_'
    – Use tokenizer algorithm in Python's Spacy to tokenize strings
    – Remove stopwords (list taken from Python's package sklearn (ENGLISH_STOP_WORDS)
    – Remove tokens strictly smaller than five characters

- Build bigram model based on text as a list of words (I use a minimum count of 5 occurrences)

- Transform text into lemmatized ngrams (using Spacy's lemmatizer)

- Build the dictionnary from lemmatized ngrams (filtering no less than in 10 documents and not more than into 60% of the corpus).

- Build LDA models from lemmatized ngrams

A1.2   Articles

The cleaning procedure for articles follow closely the one adopted for patents. However, more specific steps are required. For most articles, the full text downloaded from ScienceDirect is the result of an imperfect conversion of images into machine-encoded text: some words are not well recognized especially when the article contained mathematical symbols and equations. Words are also sometimes not properly separated by space. Additionnaly, the texts typically contain a list of references.

- Detect reference list and remove. I use a simple rule: if the word 'references' is found in the text, and if the word is located towards the end of the document (after 80% of it to be precise), I truncate the document to everything that is before. (This step is done before searching and counting molecule names).

- In addition to removing tokens that are shorter than 5 characters, I also remove tokens that are longer than 15 characters. Althgouh this simple rule may result in dropping important scientific words, it also effectively removes most of the many strings with incoherent combinations of characters.

- Drop non-English articles. Some articles seem not to be written in English. For this reason, I use Google's CLD2 library in Python to detect every document's language, and drop those that are detected with large enough confidence as not being English.

A1.3 Meta-Data

Scopus's meta-data provides the name and geographic localization of authors' affiliations. However, Scopus does not provide information about these organizations. In particular, knowing the share of articles affiliated with public vs. private entities would be interesting. To that aim, I leverage the Global Research Identifier Database[47] (GRID) which provides information about a worldwide collection of organizations associated with academic research. In particular, GRID classifies an entity as one of the following types: education, company, government, facility, non-profit, health care[48].

An organization is classified as "education" if it can grant degrees, as "company" if it is a business entity with the aim of gaining profit, as "government" if it is operated mainly by a government, and as "health care" if it is a place that treats patients. Facilities encompass building or facilities researching specific areas and usually containing specific equipment (e.g., a nuclear plant). Nonprofits include charities but also non-governmental research institutes[49].

Unfortunately, the name of the organizations and its geographical location are often reported differently in Scopus and GRID. To match as many entities as possible, I first look for exact matches, then for approximate ones using tools such as fuzzy matching in python. Still, many remained unmatched. I then manually match any organization appearing, at least, three times or more in the data. There were about 300 of such organizations.

For patents, the bulk data provided by the UPSTO contains meta-data. Names and addresses of the inventors and assignee are therefore more readily available. I use the country of the assignee, and when the patent has no assignee, I use the country of the inventor. The USPTO data, however, does not classify assignee by type of organization (e.g., company, education or non-profit). The GRID database here is not as useful because most patents originate from businesses; GRID encompasses some for-profit entities with major research activities, but many patentees are in fact small companies unlikely to be listed under GRID.

To match patent assignees to an organization type, I implement a more basic strategy. I leverage the presence of certain tokens in the name of the assignees to infer their type. For example, the "Inc." abbreviation in the name *Flow Vision, Inc.* tells us that it is a for-profit organization. Other such tokens includes "corp.", "co.", "plc", "llc", "limited" or "company", as well as "& cie"[50]. Similarly, I identify organizations containing tokens such as "university" or "school" as being of the "education" type, and those containing tokens such as "govern", "ministr" or "agency"

---

[47]https://www.grid.ac/

[48]There are two other classifications: "archive" and "other." For more information, see https://www.grid.ac/pages/policies

[49]For example, in the USA, the National Academy of Sciences is classified as a non-profit.

[50]In other languages, here are a few of the tokens that I found in the data: "kaisha" or "kk" in Japanese, "spa" in Italian, "gesellschaft" or "gmbh" or "ag" or "kg" in German, "bv" or "nv" in Dutch, "sa" or "sarl" in French, "ab" in Swedish, "oy" in Finnish, "rt" in Hungarian.

as being of the "government" type. The use of these simple rules helps me match about 36529 out of 45820 assignee names. Out of the 7899 remaining, I manually match those that appear at least ten times in my data (about 200 of them). I leave the rest with no type information.

A2 Topic Modeling
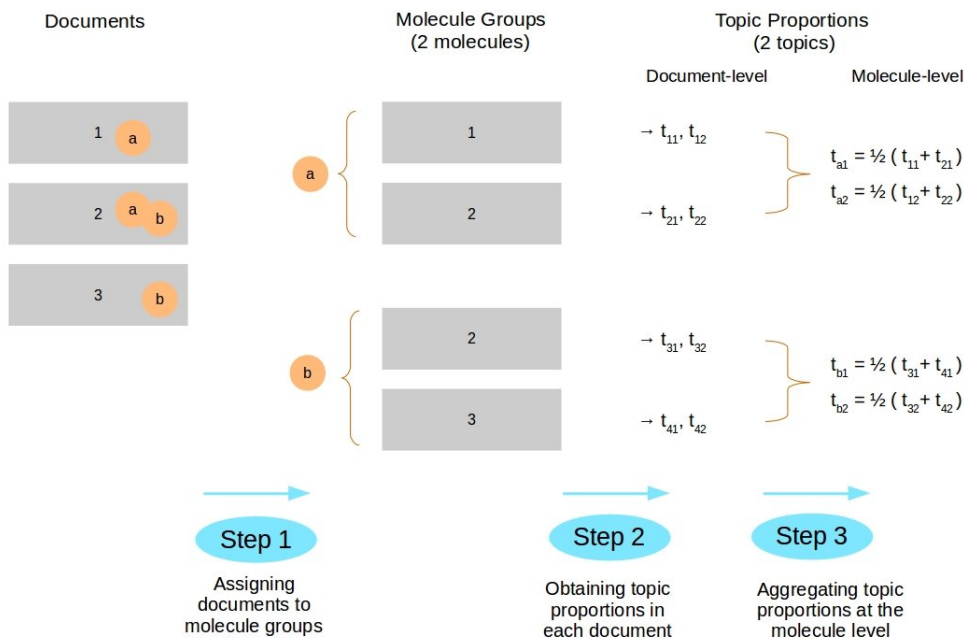
Figure A1: Schematic explanation of the methodology

Note: Suppose there are three documents: document 1 and 2 mention molecule 'a' while document 2 and 3 mention molecule 'b'. In step 1, I aggregate documents according to their molecule group. I follow a basic rule that assign any document with at least one mention of a molecule to that molecule's group. In step 2, I use topic modeling to obtain the proportions of topics in each document. $t_{i,j}$ stands for the proportion of topic j in document i. Finally, in step 3, I create a topic proportion at the molecule level by averaging over all the documents that mention the molecule of interest.
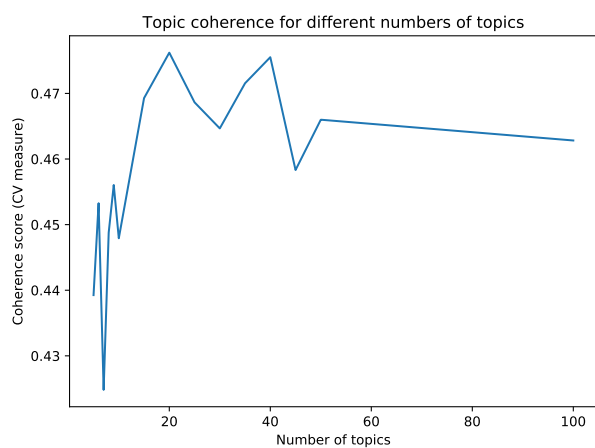
(a) Patents: 20 topics
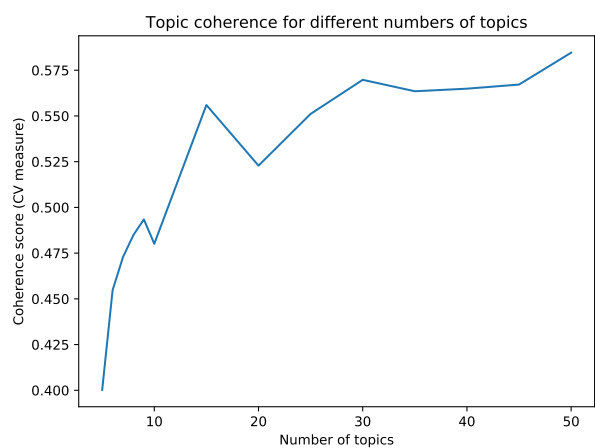
(b) Articles: 15 topics

Figure A2: Topic Coherence Scores

Table A1: Top twenty words for topics in patents

Topic 1 Words	Prob	Topic 2 Words	Prob	Topic 3 Words	Prob	Topic 4 Words	Prob	Topic 5 Words	Prob	Topic 6 Words	Prob	Topic 7 Words	Prob
polymer	0.0161	metal	0.0084	formula	0.0118	agent	0.0147	ester	0.0127	formula	0.0245	catalyst	0.0262
catalyst	0.0123	membrane	0.0082	carbon	0.0092	composition	0.0112	methyl	0.0085	atom	0.0243	metal	0.0117
carbon	0.0095	solution	0.0071	atom	0.0088	active	0.0064	titanium	0.0078	carbon	0.0214	hydrocarbon	0.0093
weight	0.0094	particle	0.0067	substitute	0.0086	weight	0.0062	catalyst	0.0060	represent	0.0175	hydrogen	0.0083
atom	0.0087	surface	0.0065	amine	0.0077	water	0.0052	solvent	0.0059	methyl	0.0128	water	0.0077
polymerization	0.0082	polymer	0.0064	metal	0.0076	solution	0.0050	ethyl	0.0056	hydrogen	0.0098	liquid	0.0074
metal	0.0065	water	0.0053	ester	0.0070	effect	0.0046	virus	0.0047	alpha	0.0094	carbon	0.0073
composition	0.0057	catalyst	0.0052	butyl	0.0070	tissue	0.0044	accord	0.0047	substitute	0.0094	component	0.0068
formula	0.0056	protein	0.0050	solvent	0.0069	formulation	0.0042	formula	0.0046	amino	0.0090	pressure	0.0068
solution	0.0056	electrode	0.0045	ether	0.0067	treatment	0.0039	agent	0.0046	radical	0.0083	oxide	0.0063
aromatic	0.0053	sample	0.0042	hydrogen	0.0066	patient	0.0039	polymerization	0.0046	general	0.0072	solvent	0.0062
prepare	0.0053	antibody	0.0039	methyl	0.0065	effective	0.0037	solution	0.0042	phenyl	0.0057	phase	0.0059
radical	0.0052	bind	0.0038	catalyst	0.0064	pharmaceutical	0.0037	active	0.0041	alkoxy	0.0056	stream	0.0057
range	0.0052	cecc	0.0038	weight	0.0060	release	0.0036	water	0.0039	halogen	0.0055	range	0.0053
component	0.0051	liquid	0.0037	phenyl	0.0060	substance	0.0036	weight	0.0039	agent	0.0054	reactor	0.0051
solvent	0.0051	enzyme	0.0036	organic	0.0058	polymer	0.0035	ether	0.0038	hydroxy	0.0054	weight	0.0049
water	0.0050	concentration	0.0035	composition	0.0056	solvent	0.0034	chloride	0.0038	derivative	0.0053	solution	0.0047
prefer	0.0047	solid	0.0033	acid	0.0053	administration	0.0034	hydrocarbon	0.0037	ethyl	0.0050	oxygen	0.0043
molecular	0.0047	electrolyte	0.0032	agent	0.0051	preparation	0.0032	solid	0.0037	solvent	0.0049	organic	0.0042
organic	0.0039	range	0.0032	radical	0.0046	ingredient	0.0031	component	0.0037	alkyl	0.0049	condition	0.0041

Topic 8 Words	Prob	Topic 9 Words	Prob	Topic 10 Words	Prob	Topic 11 Words	Prob	Topic 12 Words	Prob	Topic 13 Words	Prob	Topic 14 Words	Prob
formula	0.0288	layer	0.0265	paper	0.0145	composition	0.0170	composition	0.0127	water	0.0221	solvent	0.0185
substitute	0.0137	image	0.0200	color	0.0144	weight	0.0129	metal	0.0088	solution	0.0140	formula	0.0147
hydrogen	0.0112	silver	0.0165	pigment	0.0115	surfactant	0.0124	formula	0.0087	composition	0.0117	water	0.0078
low	0.0112	color	0.0107	solvent	0.0097	carbon	0.0096	ester	0.0079	aqueous	0.0088	methyl	0.0077
methyl	0.0095	halide	0.0105	print	0.0080	alcohol	0.0092	peptide	0.0072	metal	0.0088	solution	0.0069
phenyl	0.0088	light	0.0101	water	0.0068	water	0.0091	agent	0.0056	agent	0.0082	active	0.0068
amino	0.0078	photographic	0.0091	sheet	0.0065	agent	0.0087	carbon	0.0056	weight	0.0080	polymer	0.0063
represent	0.0075	sensitive	0.0084	agent	0.0063	atom	0.0080	acid	0.0053	particle	0.0062	ethyl	0.0056
carbon	0.0074	emulsion	0.0083	formula	0.0059	polymer	0.0067	hydrogen	0.0053	sodium	0.0062	hydrogen	0.0053
solvent	0.0072	agent	0.0081	printing	0.0058	ester	0.0066	amino	0.0051	add	0.0050	weight	0.0052
radical	0.0064	represent	0.0079	composition	0.0057	oxide	0.0065	water	0.0051	soluble	0.0045	composition	0.0052
atom	0.0063	develop	0.0063	weight	0.0053	detergent	0.0060	catalyst	0.0051	organic	0.0043	agent	0.0050
salt	0.0061	formula	0.0061	organic	0.0049	glycol	0.0059	atom	0.0050	resin	0.0042	prepare	0.0047
alkoxy	0.0061	element	0.0061	carbon	0.0047	fatty	0.0058	solution	0.0050	solid	0.0041	carry	0.0047
derivative	0.0060	coupler	0.0058	methyl	0.0047	chain	0.0051	solvent	0.0049	surface	0.0040	chloride	0.0046
prepare	0.0057	charge	0.0053	liquid	0.0045	formula	0.0051	weight	0.0047	alkali	0.0039	organic	0.0044
agent	0.0056	solution	0.0052	ester	0.0040	prefer	0.0049	prepare	0.0046	concentration	0.0038	add	0.0043
optionaccy	0.0056	developer	0.0050	microcapsule	0.0036	methyl	0.0046	radical	0.0044	oxide	0.0038	prefer	0.0043
ethyl	0.0051	substitute	0.0049	metal	0.0035	ethylene	0.0045	organic	0.0043	range	0.0037	represent	0.0042
alkyl	0.0050	photosensitive	0.0049	aqueous	0.0035	ether	0.0045	salt	0.0043	calcium	0.0036	sodium	0.0041

Topic 15 Words	Prob	Topic 16 Words	Prob	Topic 17 Words	Prob	Topic 18 Words	Prob	Topic 19 Words	Prob	Topic 20 Words	Prob
polycarbonate	0.0101	polymer	0.0229	layer	0.0227	sequence	0.0094	surface	0.0108	composition	0.0114
solution	0.0095	resin	0.0212	substrate	0.0132	cecc	0.0085	layer	0.0067	weight	0.0106
weight	0.0070	weight	0.0193	silicon	0.0099	protein	0.0075	mean	0.0062	polyester	0.0082
metal	0.0060	composition	0.0146	surface	0.0093	plant	0.0062	portion	0.0057	radical	0.0081
composition	0.0052	copolymer	0.0131	semiconductor	0.0092	amino	0.0055	member	0.0054	formula	0.0079
water	0.0050	monomer	0.0119	device	0.0091	activity	0.0053	second	0.0047	component	0.0077
alpha	0.0048	vinyl	0.0075	fiber	0.0083	growth	0.0053	sheet	0.0046	polyol	0.0075
hydroxyphenyl	0.0048	coating	0.0069	region	0.0072	enzyme	0.0052	pressure	0.0045	glycol	0.0072
acid	0.0045	agent	0.0068	oxide	0.0064	medium	0.0052	device	0.0044	isocyanate	0.0066
polymer	0.0044	polymerization	0.0061	crystal	0.0062	culture	0.0049	object	0.0043	agent	0.0065
prepare	0.0044	component	0.0060	electrode	0.0060	nucleic	0.0039	apparatus	0.0042	polymer	0.0062
atom	0.0041	rubber	0.0058	light	0.0058	microorganism	0.0038	metal	0.0039	carbon	0.0061
sodium	0.0041	acrylate	0.0057	liquid	0.0056	carbon	0.0034	fluid	0.0038	polyurethane	0.0061
catalyst	0.0040	property	0.0057	optical	0.0054	composition	0.0031	control	0.0035	atom	0.0060
methyl	0.0040	coat	0.0057	second	0.0054	prefer	0.0031	support	0.0035	catalyst	0.0059
ester	0.0039	layer	0.0056	metal	0.0054	acid	0.0031	plate	0.0034	aromatic	0.0059
solvent	0.0039	particle	0.0054	structure	0.0045	molecule	0.0031	position	0.0034	amine	0.0059
prefer	0.0038	surface	0.0054	etch	0.0044	strain	0.0030	chamber	0.0033	organic	0.0057
preparation	0.0038	solvent	0.0052	laser	0.0040	formula	0.0030	liquid	0.0032	ester	0.0056
effect	0.0037	part	0.0051	source	0.0039	peptide	0.0030	element	0.0032	molecular	0.0052

Table A2: Top twenty words for topics in articles

Topic 1 Words	Prob	Topic 2 Words	Prob	Topic 3 Words	Prob	Topic 4 Words	Prob	Topic 5 Words	Prob
compound	0.0162	surface	0.0155	laser	0.0129	gifhttps	0.0351	complex	0.0584
extract	0.0072	material	0.0096	signal	0.0102	thumbnail	0.0282	ligand	0.0261
structure	0.0068	layer	0.0086	sample	0.0097	downsample	0.0270	metal	0.0187
product	0.0061	film	0.0075	pulse	0.0092	smlhttps	0.0190	spectra	0.0141
methyl	0.0056	process	0.0062	radical	0.0081	stripin	0.0175	structure	0.0080
spectrum	0.0051	growth	0.0057	light	0.0067	yield	0.0112	coordination	0.0069
carbon	0.0051	sample	0.0050	measurement	0.0065	smlsmlimage	0.0095	tran	0.0067
japan	0.0049	particle	0.0044	intensity	0.0065	product	0.0091	spectrum	0.0067
plant	0.0049	substrate	0.0043	spectra	0.0064	gifgifaltimg	0.0090	band	0.0064
signal	0.0048	energy	0.0042	flame	0.0060	gifsisi	0.0090	compound	0.0057
aromatic	0.0048	solid	0.0040	spectrum	0.0056	compound	0.0089	coordinate	0.0055
spectra	0.0045	accoy	0.0040	absorption	0.0053	mixture	0.0089	inorg	0.0053
degradation	0.0043	pressure	0.0039	experiment	0.0052	gifgifimage	0.0088	specie	0.0051
proton	0.0042	property	0.0038	radiation	0.0051	synthesis	0.0082	stretch	0.0050
isolate	0.0040	metal	0.0037	source	0.0050	smlgrgr	0.0072	bond	0.0050
presence	0.0040	phase	0.0036	optical	0.0049	gifgrgr	0.0065	copper	0.0049
fraction	0.0040	thickness	0.0035	concentration	0.0043	scheme	0.0058	raman	0.0045
natural	0.0032	electron	0.0034	measure	0.0042	add	0.0055	solid	0.0044
yield	0.0031	structure	0.0032	irradiation	0.0041	tetrahedron	0.0055	shift	0.0044
derivative	0.0031	silicon	0.0032	range	0.0039	methyl	0.0052	chemistry	0.0042

Topic 6 Words	Prob	Topic 7 Words	Prob	Topic 8 Words	Prob	Topic 9 Words	Prob	Topic 10 Words	Prob
model	0.0144	state	0.0279	protein	0.0134	water	0.0075	protein	0.0250
energy	0.0086	energy	0.0245	amino	0.0110	plant	0.0062	activity	0.0222
function	0.0071	spectra	0.0126	peptide	0.0101	concentration	0.0060	enzyme	0.0214
phase	0.0071	electron	0.0119	acid	0.0077	sample	0.0051	bind	0.0173
equation	0.0069	fluorescence	0.0118	residue	0.0077	control	0.0051	concentration	0.0097
state	0.0067	molecule	0.0109	column	0.0066	level	0.0051	membrane	0.0083
parameter	0.0063	absorption	0.0098	chromatography	0.0063	production	0.0043	substrate	0.0078
field	0.0060	transition	0.0098	buffer	0.0057	total	0.0040	inhibitor	0.0067
calculate	0.0059	excitation	0.0083	enzyme	0.0057	organic	0.0037	receptor	0.0062
number	0.0059	transfer	0.0071	sequence	0.0055	treatment	0.0037	buffer	0.0057
constant	0.0056	spectrum	0.0070	fraction	0.0045	sediment	0.0036	inhibition	0.0051
point	0.0055	emission	0.0066	activity	0.0041	growth	0.0034	liver	0.0048
calculation	0.0053	intensity	0.0064	purification	0.0039	tissue	0.0032	assay	0.0045
order	0.0048	excited	0.0064	hydrolysis	0.0039	environmental	0.0032	biochem	0.0043
liquid	0.0045	electronic	0.0061	water	0.0038	marine	0.0029	phosphate	0.0042
large	0.0043	level	0.0061	extract	0.0035	biomass	0.0028	cytochrome	0.0039
theory	0.0041	molecular	0.0059	sample	0.0034	specie	0.0026	lipid	0.0039
measure	0.0040	orbital	0.0055	product	0.0034	research	0.0026	human	0.0039
frequency	0.0040	solvent	0.0050	sugar	0.0034	high	0.0024	presence	0.0037
interaction	0.0039	charge	0.0049	glucose	0.0034	waste	0.0024	cecc	0.0036

| Topic 11 Words	Prob	Topic 12 Words	Prob	Topic 13 Words	Prob	Topic 14 Words	Prob	Topic 15 Words	Prob
structure	0.0330	sample	0.0225	cecc	0.0292	polymer	0.0274	catalyst	0.0227
crystal	0.0148	concentration	0.0152	human	0.0102	membrane	0.0131	surface	0.0185
atom	0.0143	phase	0.0142	mutation	0.0092	water	0.0120	electrode	0.0130
compound	0.0121	column	0.0131	induce	0.0083	concentration	0.0088	oxidation	0.0107
angle	0.0109	water	0.0107	mutant	0.0074	phase	0.0082	potential	0.0092
molecule	0.0103	standard	0.0098	culture	0.0071	surface	0.0079	adsorption	0.0091
bond	0.0099	chromatogr	0.0093	strain	0.0071	chain	0.0075	carbon	0.0076
hydrogen	0.0095	determination	0.0092	cancer	0.0060	weight	0.0067	oxygen	0.0072
distance	0.0085	extraction	0.0087	assay	0.0053	particle	0.0065	hydrogen	0.0069
molecular	0.0079	separation	0.0086	expression	0.0051	molecular	0.0063	concentration	0.0065
conformation	0.0059	detection	0.0081	tumor	0.0047	sample	0.0062	catal	0.0064
structural	0.0052	liquid	0.0068	damage	0.0044	polym	0.0060	reduction	0.0064
interaction	0.0052	plasma	0.0067	sequence	0.0043	property	0.0058	metal	0.0064
energy	0.0047	chromatography	0.0066	treatment	0.0043	copolymer	0.0056	support	0.0063
chemistry	0.0047	compound	0.0059	repair	0.0042	figure	0.0053	catalytic	0.0062
length	0.0046	capiccary	0.0050	control	0.0040	solvent	0.0052	oxide	0.0061
electron	0.0045	analytical	0.0050	agent	0.0036	polymerization	0.0046	process	0.0059
carbon	0.0044	retention	0.0049	clone	0.0034	blend	0.0046	specie	0.0059
diffraction	0.0044	range	0.0045	plasmid	0.0034	monomer	0.0044	activity	0.0058
parameter	0.0044	solvent	0.0045	genetic	0.0034	surfactant	0.0043	zeolite	0.0057
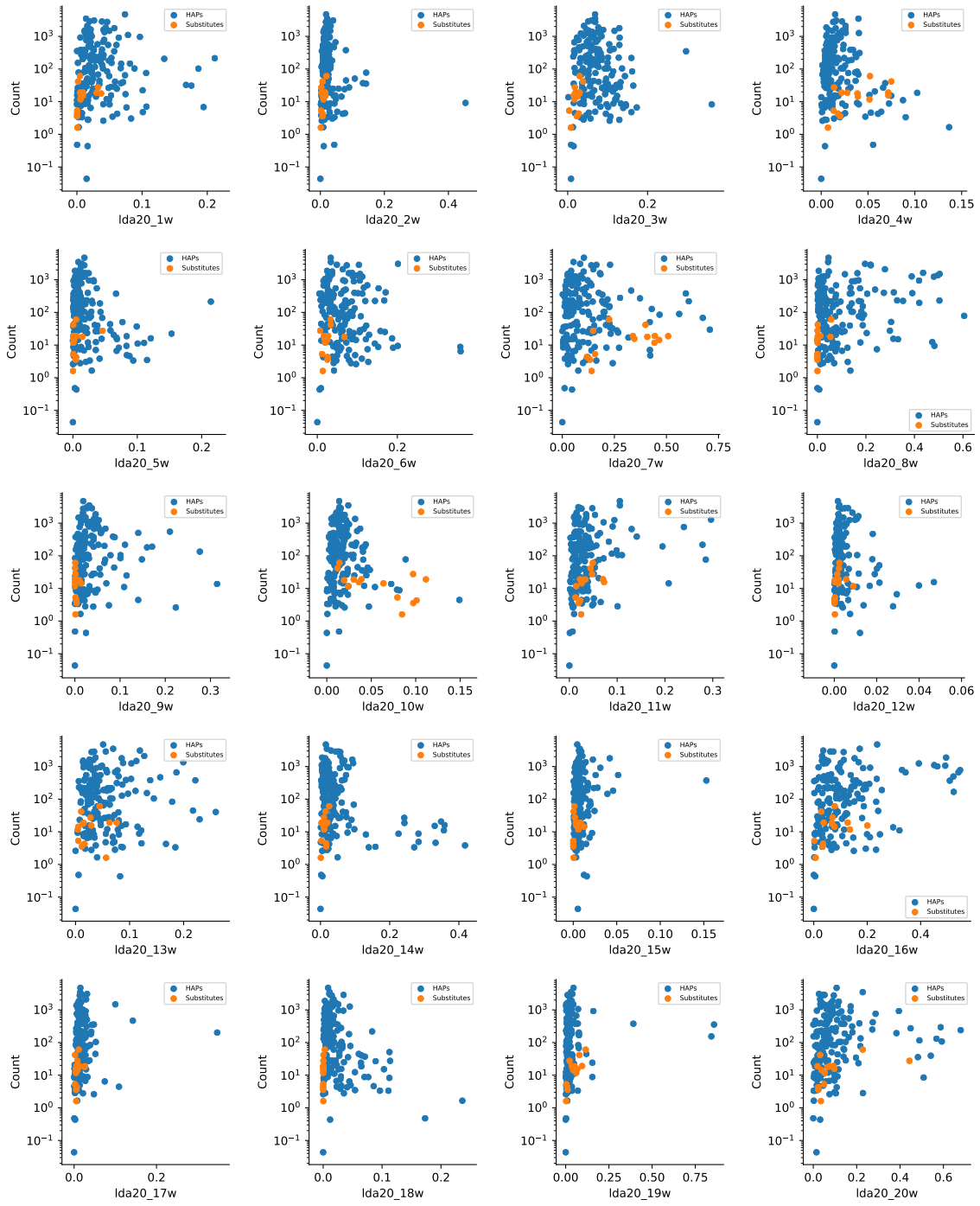
Figure A3: Scatterplot of topics proportion and count for patents.
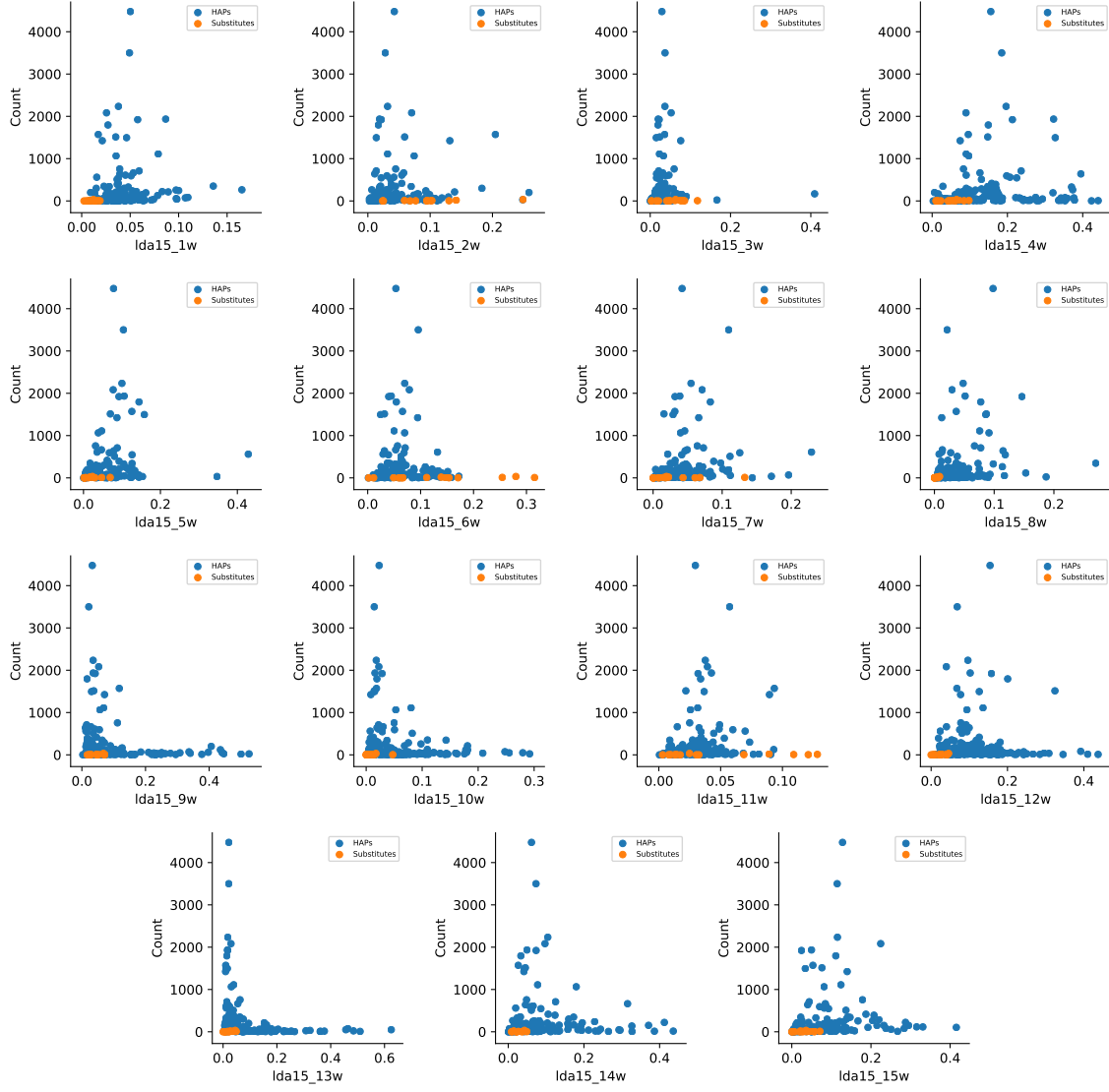
Figure A4: Scatterplot of topics proportion and count for articles.

B   Appendix B: Data and DiD Additional Figures and Tables
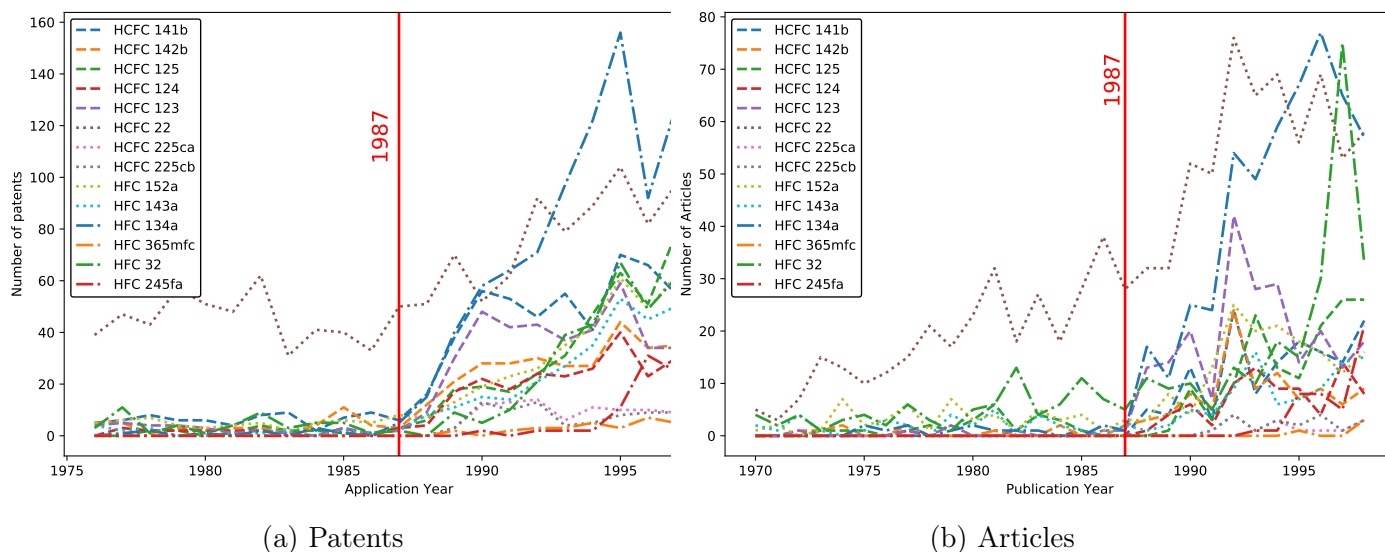

(a) Patents

(b) Articles

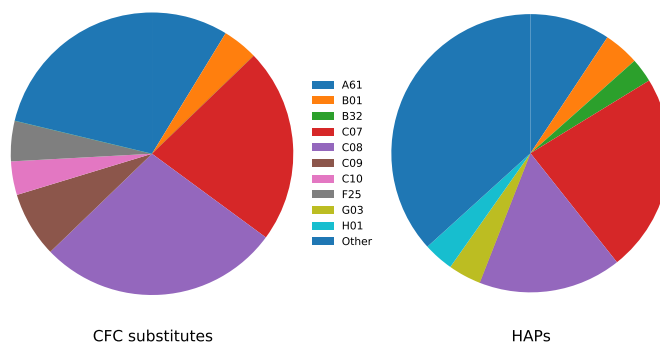Figure B1: Document Counts for Individual CFC Substitutes

Figure B2: Second-level patent codes

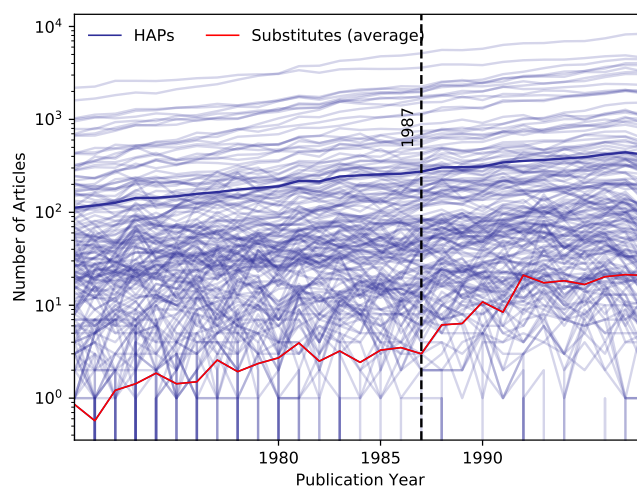Note: We see that CFC substitutes and HAPs share many second-level patent codes.

Figure B3: Articles counts for each HAP and for the average CFC substitute

Note: The grap shows article counts for each HAP (thin lines), for HAPs on average (thick line labeled "HAPs') and for CFC substitutes on average. The graph illustrates that HAPs are a diverse group of molecules. In particular, some of them have counts much higher than the average CFC substitute.

Table B1: Pre-period balance table between CFC substitutes and HAPs

(a) Patents

	HAPs	CFC substitutes	Difference	T-stat
Counts	10.88	5.36	5.52***	(4.47)
Counts (occurrence weighted)	11.75	4.19	7.56***	(5.27)
Counts (citation weighted)	15.53	9.15	6.38***	(3.44)
Counts (3-year citation weighted)	11.47	4.15	7.32***	(4.90)
Topic 1 (w. mean)	0.03	0.02	0.01	(0.98)
Topic 2 (w. mean)	0.04	0.01	0.03*	(2.56)
Topic 3 (w. mean)	0.10	0.02	0.08***	(6.91)
Topic 4 (w. mean)	0.03	0.04	-0.01	(-0.95)
Topic 5 (w. mean)	0.04	0.01	0.03**	(3.21)
Topic 6 (w. mean)	0.11	0.03	0.08***	(5.16)
Topic 7 (w. mean)	0.11	0.37	-0.26***	(-10.41)
Topic 8 (w. mean)	0.08	0.02	0.05***	(3.95)
Topic 9 (w. mean)	0.04	0.01	0.04***	(3.77)
Topic 10 (w. mean)	0.03	0.04	-0.01	(-1.16)
Topic 11 (w. mean)	0.02	0.04	-0.03***	(-3.67)
Topic 12 (w. mean)	0.01	0.01	0.00	(0.80)
Topic 13 (w. mean)	0.06	0.05	0.00	(0.06)
Topic 14 (w. mean)	0.12	0.02	0.10***	(5.41)
Topic 15 (w. mean)	0.01	0.01	-0.00	(-0.40)
Topic 16 (w. mean)	0.06	0.10	-0.03*	(-2.14)
Topic 17 (w. mean)	0.02	0.01	0.00	(0.38)
Topic 18 (w. mean)	0.04	0.00	0.03**	(3.22)
Topic 19 (w. mean)	0.02	0.07	-0.05***	(-7.30)
Topic 20 (w. mean)	0.04	0.12	-0.07***	(-4.86)

(b) Articles

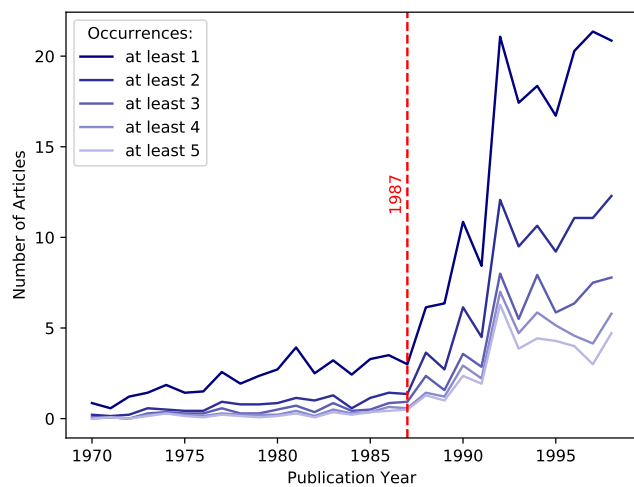|	HAPs	CFC substitutes	Difference	T-stat
Count	5.98	2.19	3.79***	(8.48)
Counts (occurrence weighted)	6.17	1.18	4.99***	(9.56)
Counts (citation weigh)	5.39	2.17	3.22***	(3.79)
Topic 1 (w. mean)	0.03	0.01	0.02***	(4.50)
Topic 2 (w. mean)	0.02	0.07	-0.04***	(-4.97)
Topic 3 (w. mean)	0.02	0.10	-0.08***	(-8.67)
Topic 4 (w. mean)	0.13	0.11	0.03	(1.36)
Topic 5 (w. mean)	0.05	0.06	-0.01	(-0.89)
Topic 6 (w. mean)	0.04	0.18	-0.13***	(-11.95)
Topic 7 (w. mean)	0.04	0.09	-0.05***	(-4.28)
Topic 8 (w. mean)	0.03	0.01	0.02***	(3.94)
Topic 9 (w. mean)	0.19	0.05	0.14***	(5.71)
Topic 10 (w. mean)	0.07	0.03	0.04***	(3.44)
Topic 11 (w. mean)	0.03	0.14	-0.11***	(-11.35)
Topic 12 (w. mean)	0.14	0.03	0.11***	(6.61)
Topic 13 (w. mean)	0.13	0.03	0.10***	(5.14)
Topic 14 (w. mean)	0.02	0.03	-0.01	(-1.19)
Topic 15 (w. mean)	0.05	0.07	-0.02*	(-2.01)

Note: The table displays the pre-period mean of outcome variables and topic proportions for patents and articles for CFC substitutes and for HAPs selected in the DiD sample.
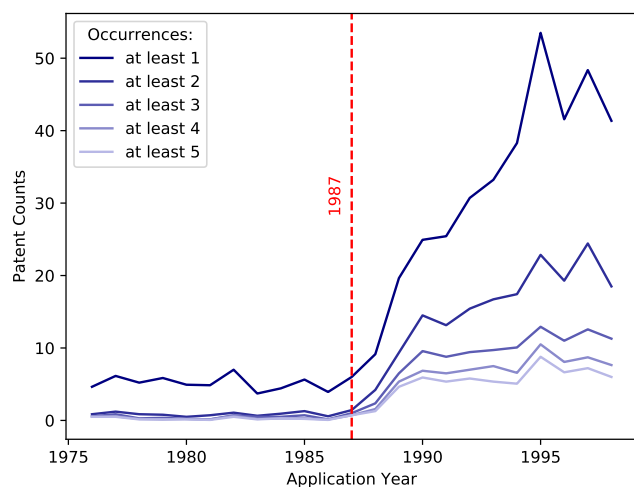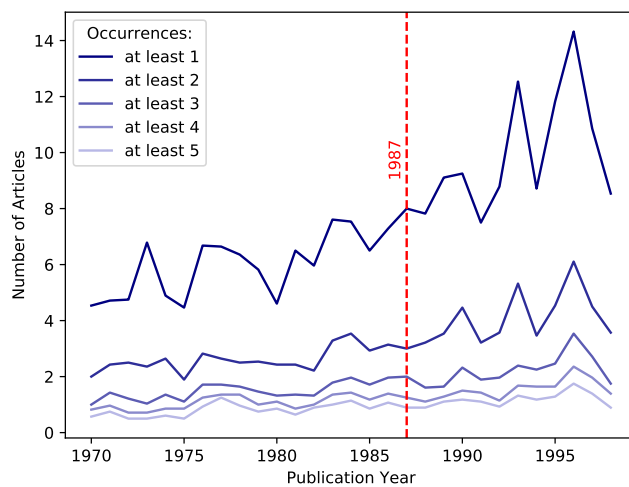
(a) Patents - CFC substitutes
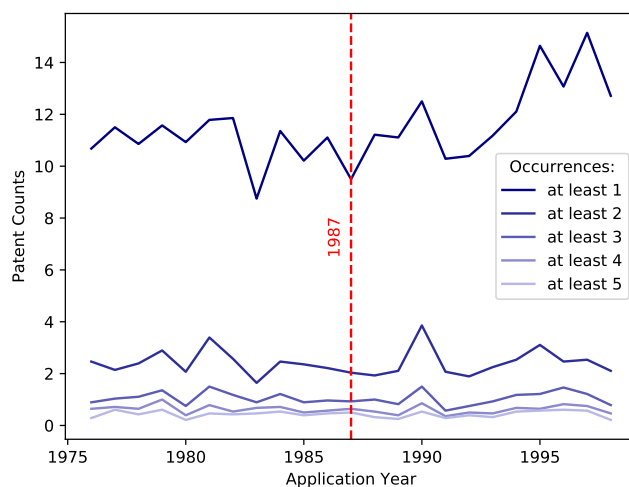
(b) Patents - HAPs

(c) Articles - CFC substitutes

(d) Articles - HAPs

Figure B4: Robustness check: counts with several thresholds of molecule occurrences

Note: The graphs illustrate that the differential trends CFC substitutes and HAPs are not affected by adopting more stringent definition of what constitutes a document "about CFC substitutes".

Table B2: Difference-in-Differences with Triadic Patents Only

	(1)	(2)
Post 1987 x Substitutes	9.473***	3.370*
	(1.190)	(1.899)
Post 1987 x Substitutes x Years		2.814***
		(0.656)
Substitutes x Years		-0.275**
		(0.121)
Years		0.583***
		(0.071)
Post 1987		-1.555**
		(0.667)
Year FE	Yes	No
Molecule FE	Yes	Yes
R-squared	0.709	0.720
Observations	714	714
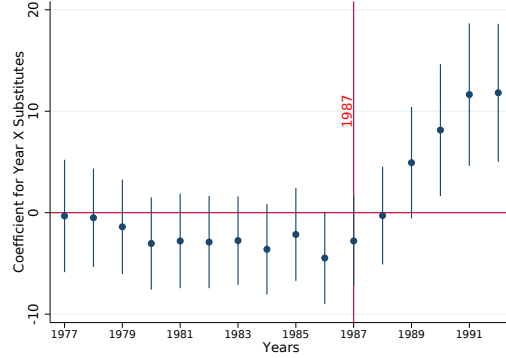
Standard errors in parentheses
Dependent variable: Number of Triadic Patents.
Years are relative to 1987.
Time span: 1976 to 1992
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
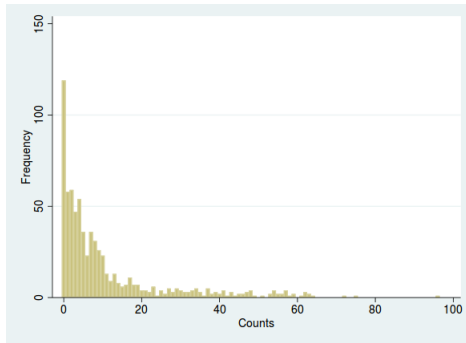
(a) Patent
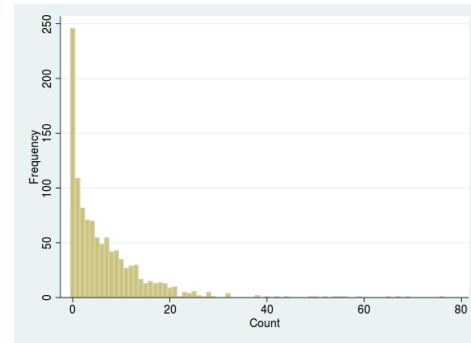
(b) Article

Figure B5: Histogram of Counts in DiD sample

Note: We see that the distribution of counts is, in both cases, zero-inflated and over-dispersed. Hence, a Zero-Inflated Negative Binomial model is preferrable to a Poisson model.

Table B3: Difference-in-differences with Zero-Inflated Negative Binomial Specifications

(a) Patents

	(1) Count	(2) Count	(3) Count	(4) Citations	(5) Occurrences	(6) Citations-Occurrences
Post 1987 x Substitutes	1.733***	1.619***	1.317***	1.614***	2.211***	2.181***
	(0.132)	(0.129)	(0.130)	(0.142)	(0.172)	(0.175)
Count (lag 1)			0.015***			
			(0.003)			
Count (lag 2)			0.008**			
			(0.004)			
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
Topics (weighted)	No	Yes	Yes	Yes	Yes	Yes
R-squared						
Observations	714	595	528	595	595	595

Zero-inflated negative binomial regression.
Dependent variable: Number of Patents.
Time span: 1976 to 1992

(b) Articles

	(1) Count	(2) Count	(3) Count	(4) Citations	(5) Occurrences	(6) Citations-Occurrences
Post 1987 x Substitutes	1.197***	0.827***	0.506***	1.407***	1.456***	2.163***
	(0.143)	(0.124)	(0.124)	(0.274)	(0.157)	(0.220)
Count (lag 1)			0.009**			
			(0.004)			
Count (lag 2)			0.012***			
			(0.004)			
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
Topics (weighted)	No	Yes	Yes	No	Yes	No
R-squared						
Observations	840	676	613	840	676	840

Zero-inflated negative binomial regression.
Dependent variable: Number of Articles.
Time span: 1976 to 1995

C   Appendix C: SCM Additional Figures and Tables

C1   SCM Theoretical Foundations

Here, I briefly summarize the theoretical underpinnings of the SCM. Suppose there are J+1 molecules, J molecules as potential controls and one, denoted with the subscript 1, that is treated. The treatment effect can be written as $\alpha_{it} = Y_{it}^T - Y_{it}^N$, where Y_{it}^N is the number of document mentioning molecule i in year t if no intervention, and Y_{it}^T the number of documents mentioning molecule i in year t if intervention. Here the quantity we need to estimate is Y_{it}^N. Abadie et al. (2010) show that a weighted average of the control units can approximate the counterfactual Y_{it}^N, that is:

$$Y_{1,t}^N \to \sum_{j=2}^{J+1} w_j^* Y_{jt} \text{ with } w* \text{ s.t. } \sum_{j=2}^{J+1} w_j^* Y_{jt} = Y_{1,t} \text{ and } \sum w_j^* Z_j = Z_1$$

To understand why this is the case, Equation 5 presents the underlying factor model. δ_t is an unknown common factor w constant loadings across units; θ_t is a vector of unknown parameters; Z_i a vector of observed covariates (not affected by intervention); λ_t unobserved common factors; μ_i a vector of unknown factor loadings and ϵ_{it} unobserved transitory shocks with zero mean. Note that this model generalizes the difference-in-differences model which imposes that λ_t be constant for all t. Hence, the unobserved confounders are constant in time and can be eliminated by taking time difference. Here, the SCM allows the effects of confounding unobserved characteristics to vary with time; taking time differences would not get us rid of μ_i.

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it} \tag{5}$$

A synthetic control such that $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ and $\sum w_j^* \mu_j = \mu_1$ would be unbiased estimator of Y_{1t}^N. In other words, fitting Z_1 and $Y_{11} \ldots Y_{1T_0}$ is a way of indirectly fitting μ_1, the unobserved factor loadings. As a result, it is important to restrict the donor pool to units with outcomes that are thought to be driven by the same structural process as for unit representing the case of interest and that were not subject to structural shocks to the outcome variable during the sample period.
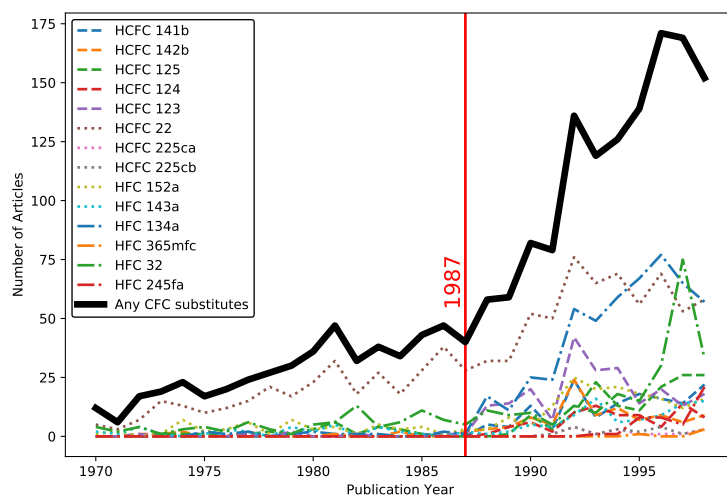
Figure C1: Article counts for CFC substitute, individually and aggregated

Note: The graph illustrates the difference between considering the 14 molecules independently and considering them as one treated molecule. The thick line called "Any CFC substitutes" corresponds to the number of articles mentioning any of the 14 CFC substitutes.

Table C1: SCM Extrapolation Check

(a) Patents

Variables (pre-1986 average)	Substitutes	HAPs Mean	HAPs Min	HAPs Max	HAPs Std.Dev.
Count	34.36	59	36.45	87.55	19.19
Topic 1 (weighted mean)	0.01	0.04	0.01	0.1	0.03
Topic 2 (weighted mean)	0.14	0.04	0	0.19	0.05
Topic 3 (weighted mean)	0.07	0.08	0.01	0.18	0.04
Topic 4 (weighted mean)	0.08	0.01	0	0.03	0.01
Topic 5 (weighted mean)	0.03	0.02	0	0.08	0.02
Topic 6 (weighted mean)	0.26	0.06	0.01	0.14	0.04
Topic 7 (weighted mean)	0.07	0.19	0.01	0.74	0.21
Topic 8 (weighted mean)	0.01	0.09	0	0.33	0.09
Topic 9 (weighted mean)	0.05	0.03	0	0.09	0.03
Topic 10 (weighted mean)	0.02	0.02	0	0.1	0.02
Topic 11 (weighted mean)	0.09	0.04	0	0.2	0.04
Topic 12 (weighted mean)	0.04	0.01	0	0.03	0.01
Topic 13 (weighted mean)	0.04	0.06	0.01	0.3	0.07
Topic 14 (weighted mean)	0.04	0.04	0.01	0.11	0.03
Topic 15 (weighted mean)	0.04	0.01	0	0.04	0.01
Topic 16 (weighted mean)	NaN	0.08	0.02	0.23	0.06
Topic 17 (weighted mean)	NaN	0.01	0	0.02	0.01
Topic 18 (weighted mean)	NaN	0.02	0	0.07	0.02
Topic 19 (weighted mean)	NaN	0.02	0	0.07	0.02
Topic 20 (weighted mean)	NaN	0.14	0.02	0.57	0.16

(b) Articles

Variables (pre-1986 average)	Substitutes	HAPs Mean	HAPs Min	HAPs Max	HAPs Std.Dev.
Count	34.36	31.38	22.27	41.82	4.85
Topic 1 (weighted mean)	0.01	0.04	0.01	0.11	0.03
Topic 2 (weighted mean)	0.14	0.03	0.01	0.07	0.02
Topic 3 (weighted mean)	0.07	0.02	0	0.1	0.02
Topic 4 (weighted mean)	0.08	0.1	0.02	0.31	0.08
Topic 5 (weighted mean)	0.03	0.04	0	0.13	0.04
Topic 6 (weighted mean)	0.26	0.05	0.01	0.18	0.05
Topic 7 (weighted mean)	0.07	0.04	0	0.24	0.05
Topic 8 (weighted mean)	0.01	0.03	0	0.08	0.02
Topic 9 (weighted mean)	0.05	0.13	0.03	0.45	0.13
Topic 10 (weighted mean)	0.02	0.08	0.01	0.25	0.07
Topic 11 (weighted mean)	0.09	0.03	0	0.08	0.02
Topic 12 (weighted mean)	0.04	0.13	0.04	0.32	0.07
Topic 13 (weighted mean)	0.04	0.16	0.01	0.49	0.15
Topic 14 (weighted mean)	0.04	0.06	0.01	0.29	0.07
Topic 15 (weighted mean)	0.04	0.05	0	0.14	0.04

Note: The table displays summary statistics for the aggregated CFC substitutes and HAPs for patents. We note that the range of values displayed by the HAPs always contains the value for CFC substitutes. Hence, the constraints that weights must sum to 1 and be non-negative does not seem to be an issue. Such constraint is imposed by the SCM algorithm to avoid extrapolation.

Table C2: Variable weights used in the construction of the synthetic control

(a) Patents

	Variable Weight
Topic 1	0.02
Topic 2	0.04
Topic 3	0.05
Topic 4	0.10
Topic 5	0.03
Topic 6	0.02
Topic 7	0.10
Topic 8	0.04
Topic 9	0.01
Topic 10	0.03
Topic 11	0.01
Topic 12	0.04
Topic 13	0.03
Topic 14	0.04
Topic 15	0.02
Topic 16	0.01
Topic 17	0.02
Topic 18	0.08
Topic 19	0.27
Topic 20	0.01
Count	0.02

(b) Articles

|	Variable Weight
Topic 1	0.06
Topic 2	0.06
Topic 3	0.07
Topic 4	0.07
Topic 5	0.06
Topic 6	0.07
Topic 7	0.02
Topic 8	0.05
Topic 9	0.02
Topic 10	0.07
Topic 11	0.13
Topic 12	0.05
Topic 13	0.12
Topic 14	0.04
Topic 15	0.07
Count	0.05

Note: The table displays the value of each variable's contribution to the synthetic control. We note that topic 19, 4 and 7 contribute the most for patents, and topic 11 and 13 for articles. This indicate that these topics had the highest correlations with the outcome variable. In the Stata *synth* package, these weights are determined according to the amount of predictive power that each variable has over the outcome.
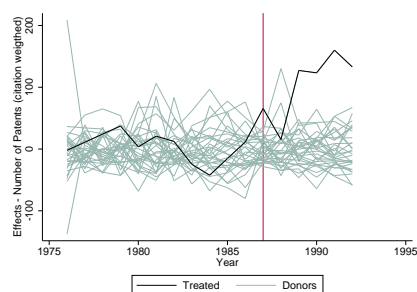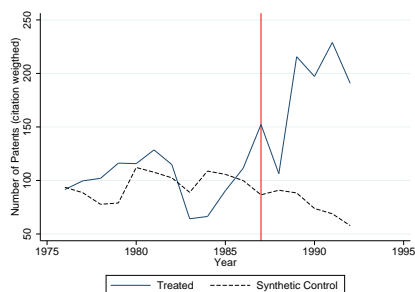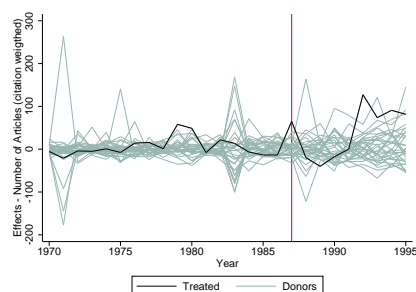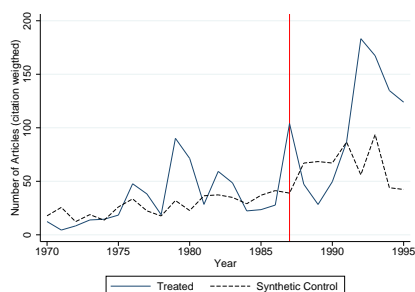
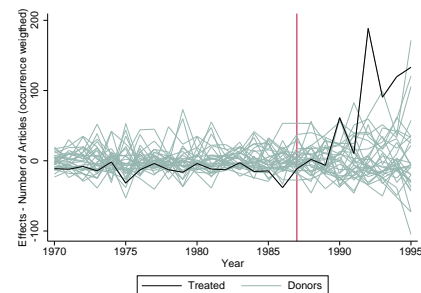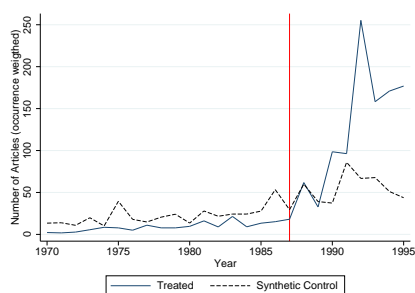(a) Citation weighted
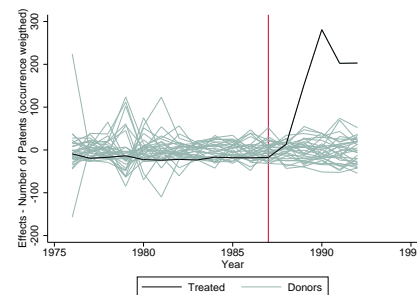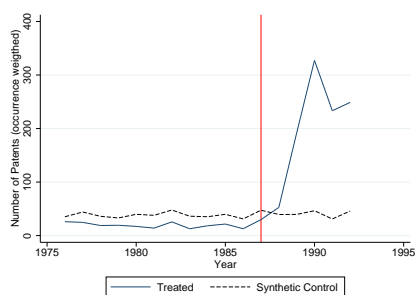
(b) Occurrence weighted
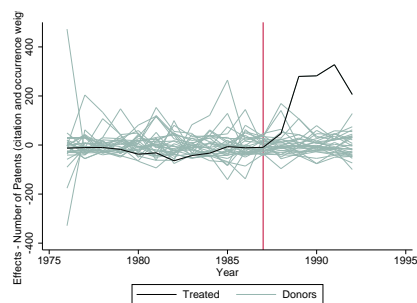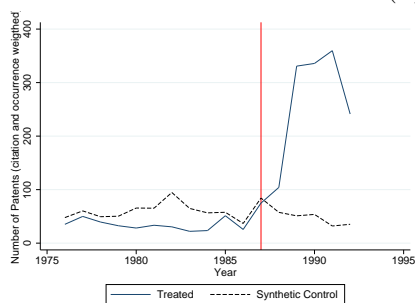
(c) Occurrence and Citation weighted

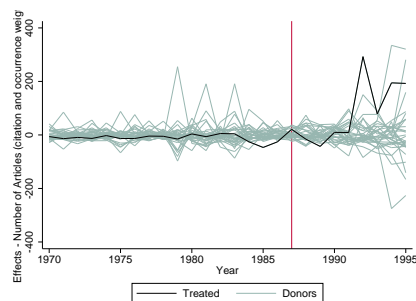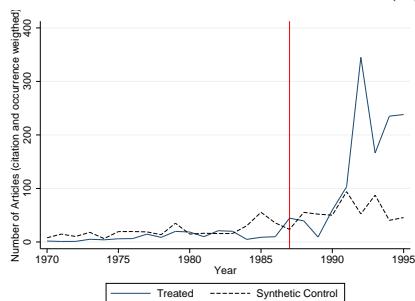Figure C2: Robustness check for Patents: SCM with counts weighted by occurrences and citations

Note: These figures show that implementing the SCM using patent counts weighted by molecule occurences and patent citation does not alter the main conclusions.
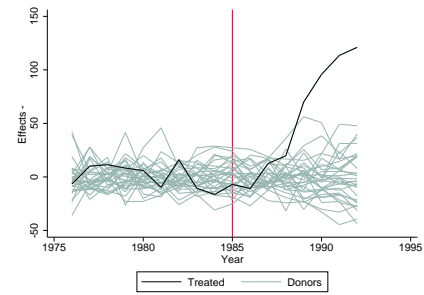
(a) Citation weighted

(b) Occurrence weighted

(c) Occurrence and Citation weighted

Figure C3: Robustness check for Patents: SCM with counts weighted by occurrences and citations

Note: These figures show that implementing the SCM using article counts weighted by molecule occurences and article citation does not alter the main conclusions.
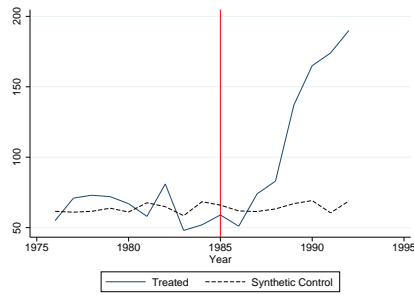
(a) Patents: raw effect (left) and placebo tests (right)
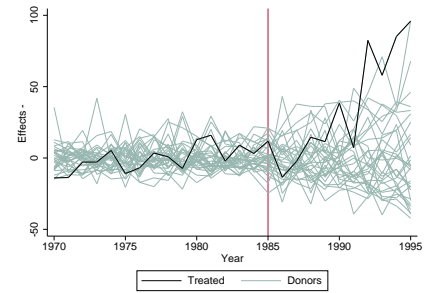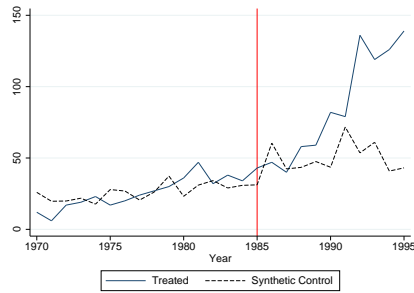
(b) Articles: raw effect (left) and placebo tests (right)

Figure C4: SCM graphs for CFC substitutes assuming anticipation

Note: These figures show that implementing the SCM using years only up to 1982 does not alter the main conclusions.

D   Appendix D: Others Figures and Tables

Table D1: Summary Statistics for Documents Mentioning CFC substitutes

(a) Patents

	count	mean	sd	min	max
Occurrences	3437	6.17	11.32	1.00	187.00
Citations	3273	9.25	13.23	0.00	153.00
USA	3179	0.59	0.49	0.00	1.00
UK	3179	0.05	0.22	0.00	1.00
Japan	3179	0.19	0.39	0.00	1.00
Canada	3179	0.00	0.07	0.00	1.00
France	3179	0.03	0.17	0.00	1.00
Germany	3179	0.09	0.28	0.00	1.00
Italy	3179	0.01	0.11	0.00	1.00
Europe	3179	0.21	0.41	0.00	1.00
Education	3140	0.03	0.16	0.00	1.00
Company	3140	0.96	0.19	0.00	1.00
Government	3140	0.00	0.07	0.00	1.00
Facilities	3140	0.00	0.07	0.00	1.00
Non Profit	3140	0.00	0.00	0.00	0.00
Healthcare	3140	0.00	0.00	0.00	0.00

(b) Articles

	count	mean	sd	min	max
Occurrences	1926	7.18	16.53	1.00	222.00
Citations	926	31.74	70.58	0.00	1298.00
USA	892	0.37	0.48	0.00	1.00
Japan	892	0.09	0.29	0.00	1.00
UK	892	0.10	0.31	0.00	1.00
Germany	892	0.08	0.28	0.00	1.00
France	892	0.05	0.22	0.00	1.00
Italy	892	0.05	0.22	0.00	1.00
Canada	892	0.05	0.22	0.00	1.00
India	892	0.03	0.17	0.00	1.00
Netherlands	892	0.04	0.19	0.00	1.00
Spain	892	0.01	0.11	0.00	1.00
Europe	892	0.38	0.49	0.00	1.00
Education	893	0.68	0.47	0.00	1.00
Company	893	0.13	0.34	0.00	1.00
Government	893	0.09	0.29	0.00	1.00
Facilities	893	0.15	0.36	0.00	1.00
Non Profit	893	0.04	0.19	0.00	1.00
Healthcare	893	0.02	0.14	0.00	1.00

Note: "Occurrences" capture the number of time any relevant molecule is mentioned in the document. "Facilities" encompass building or facilities researching specific areas and usually containing specific equipment (e.g., a nuclear plant). "Healthcare" corresponds to institutions were patients are treated (e.g. hospitals). See Section A1.3 for more details about country and affiliation data.
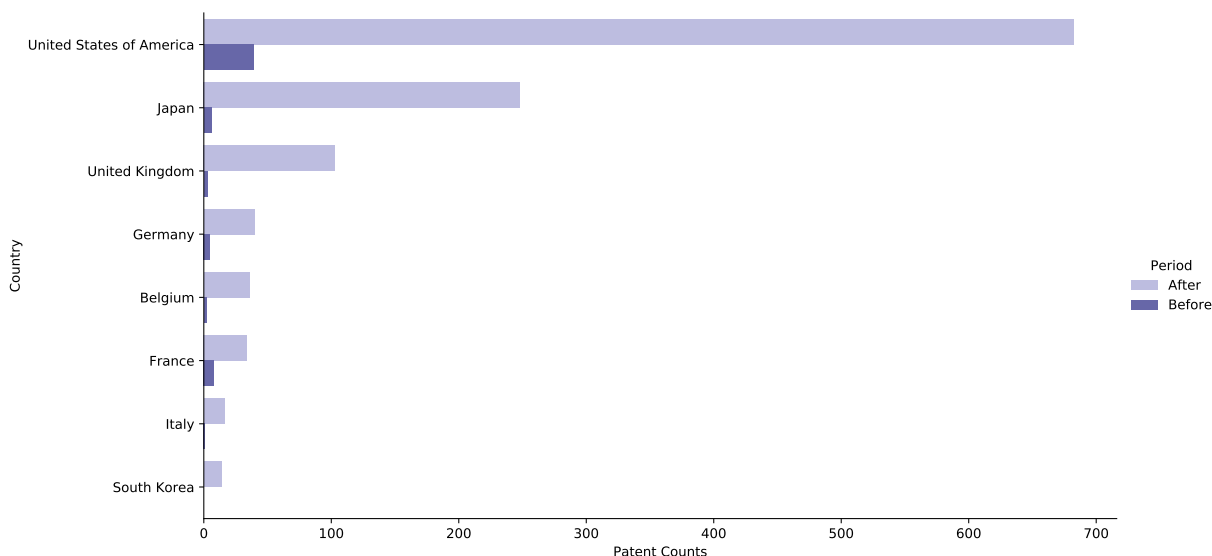
Figure D1: Patent Counts by Country before and after 1987

Table D2: Summary Statistics for Documents Mentioning CFC Substitutes Before and After 1987

(a) Patents

	Before	After	Difference	T-stat
Occurrences	1.87	7.66	-5.80***	(-13.46)
Citations	14.99	7.59	7.40***	(13.74)
USA	0.59	0.59	0.00	(0.09)
UK	0.02	0.06	-0.04***	(-4.46)
Japan	0.12	0.21	-0.09***	(-5.55)
Canada	0.01	0.00	0.00	(0.95)
France	0.04	0.03	0.01	(1.30)
Germany	0.19	0.05	0.14***	(12.12)
Italy	0.01	0.02	-0.01*	(-2.24)
Europe	0.27	0.19	0.08***	(4.78)
Education	0.02	0.03	-0.01	(-1.86)
Company	0.97	0.96	0.01	(0.77)
Government	0.01	0.00	0.01***	(4.41)
Facilities	0.00	0.01	-0.01*	(-2.14)
Non Profit	0.00	0.00	0.00	(.)
Healthcare	0.00	0.00	0.00	(.)

(b) Articles

	Before	After	Difference	T-stat
Occurrences	2.41	8.91	-6.50***	(-7.74)
Citations	31.80	31.71	0.09	(0.02)
USA	0.43	0.34	0.09**	(2.69)
Japan	0.04	0.12	-0.08***	(-3.99)
UK	0.13	0.09	0.04	(1.84)
Germany	0.06	0.10	-0.04*	(-2.17)
France	0.07	0.04	0.02	(1.58)
Italy	0.02	0.06	-0.04**	(-2.66)
Canada	0.08	0.03	0.05***	(3.55)
India	0.04	0.02	0.01	(1.02)
Netherlands	0.04	0.04	-0.01	(-0.38)
Spain	0.00	0.02	-0.02*	(-2.31)
Europe	0.37	0.39	-0.02	(-0.53)
Education	0.67	0.69	-0.02	(-0.61)
Company	0.10	0.15	-0.05	(-1.91)
Government	0.06	0.10	-0.04*	(-2.10)
Facilities	0.17	0.15	0.02	(0.89)
Non Profit	0.05	0.03	0.02	(1.22)
Healthcare	0.02	0.02	0.01	(0.60)

Note: "Occurrences" capture the number of time any relevant molecule is mentioned in the document. "Facilities" encompass building or facilities researching specific areas and usually containing specific equipment (e.g., a nuclear plant). "Healthcare" corresponds to institutions were patients are treated (e.g. hospitals). See Section A1.3 for more details about country and affiliation data.
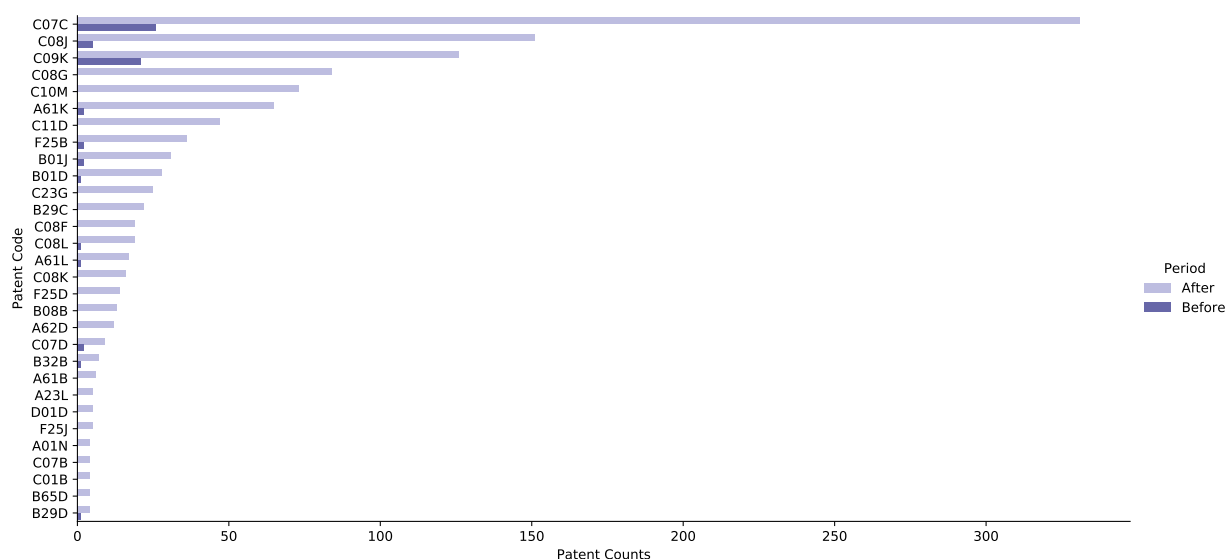
Figure D2: Most frequent codes for patents mentioning CFC susbtitutes before and after 1987

Note: The figure illustrates the differences between the most frequent codes for patents before and after 1987. The most frequent patent codes before 1987 tend to be the most frequent after 1987. At the same time, some codes with low to zero frequency before 1987 become important after 1987 (e.g., C08G, C10M, C23G or C11D). Only patents with at least 3 molecule occurrences are kept in the sample.