INDUCED INNOVATION AND INTERNATIONAL ENVIRONMENTAL AGREEMENTS: EVIDENCE FROM THE OZONE REGIME
ONLINE SUPPLEMENTARY MATERIAL

EUGENIE DUGOUA*

April 14, 2021

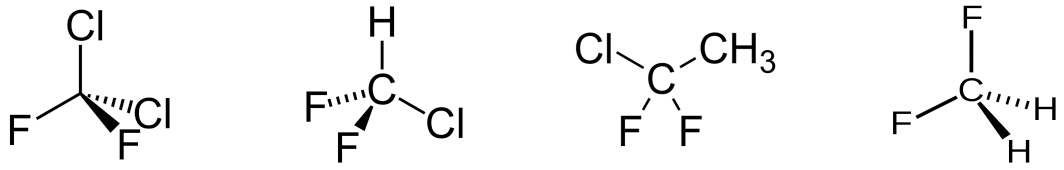
Contents

List of Figures

*Dugoua: London School of Economics, Houghton St, Holborn, London WC2A 2AE. Email: e.dugoua@lse.ac.uk.

List of Tables

A Other Useful Background Information



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

Figure A1: Molecular Structure of CFCs, HCFCs and HFCs

Note: CFC stands for chlorofluorocarbon, i.e., a molecule entirely made of carbon, chlorine, and fluorine atoms. When a hydrogen atom substitutes a chlorine atom in CFC-12, we get HCFC-22, or when, instead, a methyl group substitutes a chlorine atom, we obtain HCFC-142b. Here "HCFC" stands for hydro-chlorofluorocarbons. When hydrogens substitute all the chlorine atoms, the compounds are known as hydro-fluorocarbons (HFCs). For example, when hydrogens replace the two chlorine atoms in CFC-12, we get HFC-32.

Table A1: Montreal Protocol Phaseout Schedules

Chemicals	1987 Montreal Protocol	1990 London Revisions	1992 Copenhagen Revisions	1995 Vienna Revisions	1995 Vienna (article 5)
Annex A/I Chlorofluorocarbons 11,12,113,114,115	baseline 1986 freeze 1989 20% 1993 50% 1998	baseline 1986 freeze 1989 50% 1995 85% 1997 ...	baseline 1986 freeze 1989 75% 1994 100% 1996	no change	baseline 1995/97 freeze 1999 50% 2005 85% ...
Annex A/II Halons 1211, 1301, 2402	baseline 1986 freeze 1992	baseline 1986 freeze 1992 50% 1995 100% 2000	baseline 1986 freeze 1992 100% 1994	no change	baseline 1995/97 freeze 2002 50% 2005 100% ...
Annex B/I Other CFCs 10 chemicals	no controls	baseline 1989 20% 1993 85% 1997 100% 2000	baseline 1989 20% 1993 75% 1994 100% 1996	no change	baseline 1998/2000 20% 2003 85% 2007 100%...
Annex B/II Carbon tetrachloride		baseline 1989 85% 1995 100% 2000	baseline 1989 85% 1995 100% 1996	no change	baseline 1998/2000 85% 2005 100% 2010
Annex B/III Methyl chloroform		baseline 1989 freeze 1993 30% 1995 70% 2000 ...	baseline 1989 freeze 1993 50% 1994 100% 1996	no change	baseline 1998/2000 freeze 2013 30% 2005 70% ...
Annex C/I Hydrochlorofluorocarbons 40 chemicals	no controls	mandatory re-porting nonbiding reso-lution on pase-out: 2020 if pos...	baseline 1989 freeze 1996 35% 2004 65% 2010 90% 201...	baseline 1989 one change	baseline 2015 freeze 2016 100% 2040
Annex C/II Hydrobromofluorocarbons 34 chemicals	no controls	no controls	100% 1996	no change	100% 1996
Annex E Methyl bromide	no controls	no controls	baseline 1991 freeze 1995	baseline 1991 freeze 1995 25% 2001 50% 2005 100% 2010	baseline 1995/98 freeze 2002

Note: Source: Benedick (2009)

Table A2: List Molecules in Each Treatment Group

CFC Substitutes	HCFC 22, HCFC 123, HCFC 124, HCFC 125, HCFC 141b, HCFC 142b, HCFC 225ca, HCFC 225cb, HFC 134a, HFC 143a, HFC 152a, HFC 245fa, HFC 32, HFC 365mfc
Annex A	CFC 11, CFC 12, CFC 113, CFC 114, CFC 115, HALON 1211, HALON 1301, HALON 2402
Annex B	CFC 13, CFC 111, CFC 112, CFC 211, CFC 212, CFC 213, CFC 214, CFC 215, CFC 216, CFC 217, Carbon tetrachloride, Methyl chloroform
HAPs	Acetaldehyde, Acetamide, Acetonitrile, Acetophenone, 2-Acetylaminofluorene, Acrolein, Acrylamide, Acrylic acid, Acrylonitrile, Allyl chloride, 4-Aminobiphenyl, Aniline, o-Anisidine, Asbestos, Benzene, Benzidine, Benzotrichloride, Benzyl chloride, Biphenyl, Bis(2-ethylhexyl)phthalate (DEHP), Bis(chloromethyl)ether, Bromoform, 1,3-Butadiene, Calcium cyanamide, Caprolactam, Captan, Carbaryl, Carbon disulfide, Carbonyl sulfide, Catechol, Chloramben, Chlordane, Chlorine, Chloroacetic acid, 2-Chloroacetophenone, Chlorobenzene, Chlorobenzilate, Chloroform, Chloromethyl methyl ether, Chloroprene, Cresols/Cresylic acid, o-Cresol, m-Cresol, p-Cresol, Cumene, 2,4-D, salts and esters, DDE, Diazomethane, Dibenzofurans, 1,2-Dibromo-3-chloropropane, Dibutylphthalate, 1,4-Dichlorobenzene, 3,3-Dichlorobenzidene, Dichloroethyl ether ether), 1,3-Dichloropropene, Dichlorvos, Diethanolamine, N,N-Dimethylaniline, Diethyl sulfate, 3,3-Dimethoxybenzidine, Dimethyl aminoazobenzene, 3,3'-Dimethyl benzidine, Dimethyl carbamoyl chloride, Dimethyl formamide, 1,1-Dimethyl hydrazine, Dimethyl phthalate, Dimethyl sulfate, 4,6-Dinitro-o-cresol, and salts, 2,4-Dinitrophenol, 2,4-Dinitrotoluene, 1,4-Dioxane, 1,2-Diphenylhydrazine, Epichlorohydrin, 1,2-Epoxybutane, Ethyl acrylate, Ethyl benzene, Ethyl carbamate, Ethyl chloride, Ethylene dibromide, Ethylene dichloride, Ethylene glycol, Ethylene imine, Ethylene oxide, Ethylene thiourea, Ethylidene dichloride, Formaldehyde, Heptachlor, Hexachlorobenzene, Hexachlorobutadiene, Hexachlorocyclopentadiene, Hexachloroethane, Hexamethylene-1,6-diisocyanate, Hexamethylphosphoramide, Hexane, Hydrazine, Hydrochloric acid, Hydrogen fluoride, Hydrogen sulfide, Hydroquinone, Isophorone, Lindane, Maleic anhydride, Methanol, Methoxychlor, Methyl bromide, Methyl chloride, Methyl ethyl ketone, Methyl hydrazine, Methyl iodide, Methyl isobutyl ketone, Methyl isocyanate, Methyl methacrylate, Methyl tert butyl ether, 4,4-Methylene bis(2-chloroaniline), Methylene chloride, Methylene diphenyl diisocyanate, 4,4'-Methylenedianiline, Naphthalene, Nitrobenzene, 4-Nitrobiphenyl, 4-Nitrophenol, 2-Nitropropane, N-Nitroso-N-methylurea, N-Nitrosodimethylamine, N-Nitrosomorpholine, Parathion, Pentachloronitrobenzene, Pentachlorophenol, Phenol, p-Phenylenediamine, Phosgene, Phosphine, Phosphorus, Phthalic anhydride, Polychlorinated biphenyls, 1,3-Propane sultone, beta-Propiolactone, Propionaldehyde, Propoxur, Propylene dichloride, Propylene oxide, 1,2-Propylenimine, Quinoline, Quinone, Styrene, Styrene oxide, 2,3,7,8-Tetrachlorodibenzo-p-dioxin, 1,1,2,2-Tetrachloroethane, Tetrachloroethylene, Titanium tetrachloride, Toluene, 2,4-Toluene diamine, 2,4-Toluene diisocyanate, o-Toluidine, Toxaphene, 1,2,4-Trichlorobenzene, 1,1,2-Trichloroethane, Trichloroethylene, 2,4,5-Trichlorophenol, 2,4,6-Trichlorophenol, Triethylamine, Trifluralin, 2,2,4-Trimethylpentane, Vinyl acetate, Vinyl bromide, Vinyl chloride, Vinylidene chloride, Xylenes, o-Xylenes, m-Xylenes, p-Xylenes

Table A3: Details about CFC Substitutes

Substitute	PAFT	AFEAS	Substitute for	Notes
HCFC-22	No, already marketed, toxicology known	Yes	Included in Annex C. CFC-11, CFC-12 in foams	cheapest, fastest substitute, already at large scale production at the end of 1986 but due to toxicity concerns, not appropriate for aerosol use. FDA approved it for foams in 1988 for fast foods and for grocery display packaging.
HCFC-142b	No, already marketed, toxicology known	Yes	CFC-11, CFC-12 but not ideal	Included in Annex C. Considered because already at small scale production in 1986 but their thermodynamic propertiesare very different and would have required changes in equipment and process. DuPont 1988 process for coproduction of HCFC 141b and 142b
HFC-152a	No, already marketed, toxicology known	Yes	CFC-11, CFC-12 but not ideal	Considered because already at small scale production in 1986 but their thermodynamic propertiesare very different and would have required changes in equipment and process.
HCFC-123	Yes	Yes	CFC-11 in refrigeration	Included in Annex C. Vapor pressure similar to CFC-11 and CFC-12 implied no need to change equipment. However no commercial experience. estimated at $1.5-2/lb in 1986. DuPont patent commercial synthesis route 1988. large plant in 1990 for production. Still some toxicity concerns.
HFC-134a	Yes	Yes	CFC-12 in refrigeration (car AC)	vapor pressure similar to CFC-11 and CFC-12 implied no need to change equipment. However no commercial experience. estimated at $3/lb in 1986. oct 1990 first commercial plant ICI, then DuPont. Both DuPont and ICI announced important catalyst breakthroughs in 1992, which roughly doubled their capacity.
HCFC-141b	Yes	Yes	CFC-11 in foams	Included in Annex C. Vapor pressure similar to CFC-11 and CFC-12 implied no need to change equipment. However no commercial experience. DuPont 1988 process for coproduction of HCFC 141b and 142b. Appeared to be the most promising alternative initially (1987-1988) but in late 1988 its ODP was found much higher than thought (about 10 percent). EPA banned its use as a solvent in 1993. required phase out of production by 2003. Moderate inflammability.
HCFC-124	Yes	Yes	CFC-114 in refrigeration and sterilization	Included in Annex C. Less suitable properties but could be used in blends
HCFC-125	Yes	Yes	CFC-115 in refrigeration and sterilization	less suitable properties but could be used in blends
HCFC-225ca	No, second rank candidate	Yes		Included in Annex C.
HCFC-225cb	No, second rank candidate	Yes		Included in Annex C.
HFC-32	No, second rank candidate	Yes	refrigeration	considered in blends for refrigeration. Inflammability and compressor discharge made it problematic alone. Both DuPont and ICI opened HFC-32 plants in the summer of 1992. by 1993, DuPont, Allied, ICI, and Atochem were all marketing various patented refrigerant blends
HFC-143a	No, second rank candidate	Yes	CFC-12 in refrigeration	less suitable properties but could be used in blends
HFC-245fa	No	No	CFC-11, HCFC-141b and HCFC-142b in foams	
HFC-365mfc	No	No	CFC-11, HCFC-141b and HCFC-142b in foams	

Note: Information collected from (Parson 2003) and (Benedick 2009). Note: the cost of CFC-12 in 1986 was $0.65/lb.

Table A4: List of Substitutes and Their Possible Names

HCFC 22
Chlorodifluoromethane
Algeon 22
Algofrene 22
Algofrene 6
Arcton 22
Arcton 4
CFC 22
Daiflon 22
Difluorochloromethane
Difluoromethyl chloride
Difluoromonochloromethane
Dymel 22
Electro-CF 22
F 22 (halocarbon)
FC 22
FC 22 (halocarbon)
FKW 22
Flugene 22
Forane 22
Freon 22
Freon R 22
Frigen 22
Fron 22
Genetron 22
HFA 22
Halon 22
Haltron 22
Isceon 22
Isotron 22
Khladon 22
Korfron 22
Monochlorodifluoromethane
Propellant 22
R 22
Refrigerant 22
Refrigerant R 22
Solkane 22
Ucon 22

HCFC 123
2,2-Dichloro-1,1,1-trifluoroethane
1,1,1-Trifluoro-2,2-dichloroethane
1,1,1-Trifluorodichloroethane
1,1-Dichloro-2,2,2-trifluoroethane
CFC 123
Dichloro(trifluoromethyl)methane
F 123
F 123 (halocarbon)
FC 123
Freon 123
Fron 123
HFA 123
Khladon 123
R 123
Solkane 123

HCFC 124
2-Chloro-1,1,1,2-tetrafluoroethane
1,1,1,2-Tetrafluoro-2-chloroethane
1,1,1,2-Tetrafluorochloroethane
1-Chloro-1,2,2,2-tetrafluoroethane
CFC 124
F 124
F 124 (halocarbon)
FC 124
Freon 124
Fron 124
Khladon 124
R 124

HCFC 125
Ethane, pentafluoro- (6CI,7CI,8CI,9CI)
1,1,1,2,2-Pentafluoroethane
1,1,2,2,2-Pentafluoroethane
Ecolo Ace 125
F 125
FC 125
Freon 125
Fron 125
HFA 125
HFC 125
HFO 125
Khladon 125
Pentafluoroethane
R 125

HCFC 141b
1,1-Dichloro-1-fluoroethane
1-Fluoro-1,1-dichloroethane
141B
Asahiklin AK 141b
CFC 141b
CG 141b
Daiflon 141b
Dichlorofluoroethane
F 141b
Forane 141b
Forane DGX
Fron 141b
Genesolv 2000
Genetron 141b
HFA 141b
HFC 141b
Isotron 141b
Khladon 141b
R 141b
RC 14
Refrigerant 141b
Solkane 141b

HCFC 142b
1-Chloro-1,1-difluoroethane
1,1-Difluoro-1-chloroethane
CFC 142b
Daiflon 142b
Dymel 142
F 142b
FC 142b
FKW 142b
Freon 142b
Fron 142b
Genetron 101
Genetron 142b
HFA 142b
Propellant 142B
R 142b
Solkane 142b
α-Chloroethylidene fluoride

HCFC 152a
1,1-Difluoroethane
Algofrene 67
Dymel 152
Dymel 152A
Ethylidene fluoride
F 152A
FC 152a
FKW 152a
Formacel Z 2
Fron 152a
Genetron 152A
HFA 152a
HFC 152a
HFO 152a
Propellant 152A
R 152a
Solkane 152a TG 152a

HCFC-225ca
3,3-Dichloro-1,1,1,2,2-pentafluoropropane
1,1,1,2,2-Pentafluoro-3,3-dichloropropane
1,1-Dichloro-2,2,3,3,3-pentafluoropropane
Fron 225
R 225b
R 225ca

HCFC-225cb
1,3-Dichloro-1,1,2,2,3-pentafluoropropane
1,1,2,2,3-Pentafluoro-1,3-dichloropropane
AK 225G
AK 225cb
Asahiklin AK 225G
HFC 225bc
R 225a
R 225cb

HCFC 134a
1,1,1,2-Tetrafluoroethane
1,2,2,2-Tetrafluoroethane
AK 134a
Arcton 134a
Ecolo Ace 134a
F 134A
FC 134a
Forane 134a
Freon 134a
Fron 134a
Genetron 134a
HC 134a
HFA 134
HFA 134a
HFA P134a
HFC 134a
Halon 134A
KLEA 134a
Khladon 134a
Meforex 134a
Norflurane
P 134A
R 134a
RF 134a
Refrigerant R 134a
SUVA 134a
Solkane 134a
TG 134a

HCFC 143a
1,1,1-Trifluoroethane
CFC 143A
F 143A
FC 143a
Freon 143a
Fron 143a
HCF 143a
HFA 143a
HFC 143a
HFO 143a
Methylfluoroform
R 143a
TG 143a

HFC 245fa
1,1,1,3,3-Pentafluoropropane
1,1,3,3,3-Pentafluoropropane
245fa
Enovate 245
Enovate 245fa
Enovate 3000
Genetron 245fa

HFC 32
Difluoromethane
Ecolo Ace 32
F 32
FC 32
Forane 32
Freon 32
Genetron 32
HFA 32
HFO 32
Methylene difluoride
R 32
R 32 (refrigerant)

HFC 365mfc
1,1,1,3,3-Pentafluorobutane
2,2,4,4,4-Pentafluorobutane

Forane 365mfc
HFC 365
HFO 365mfc
R 365
R 365mfc
Solkane 365
Solkane 365mfc

B Cleaning Procedures and Topic Modelling

B1 Cleaning procedure

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

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.

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[1] (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[2].

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[3].

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

1. `https://www.grid.ac/`

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

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

"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"[4]. 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" 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.

B2 Topic Modeling

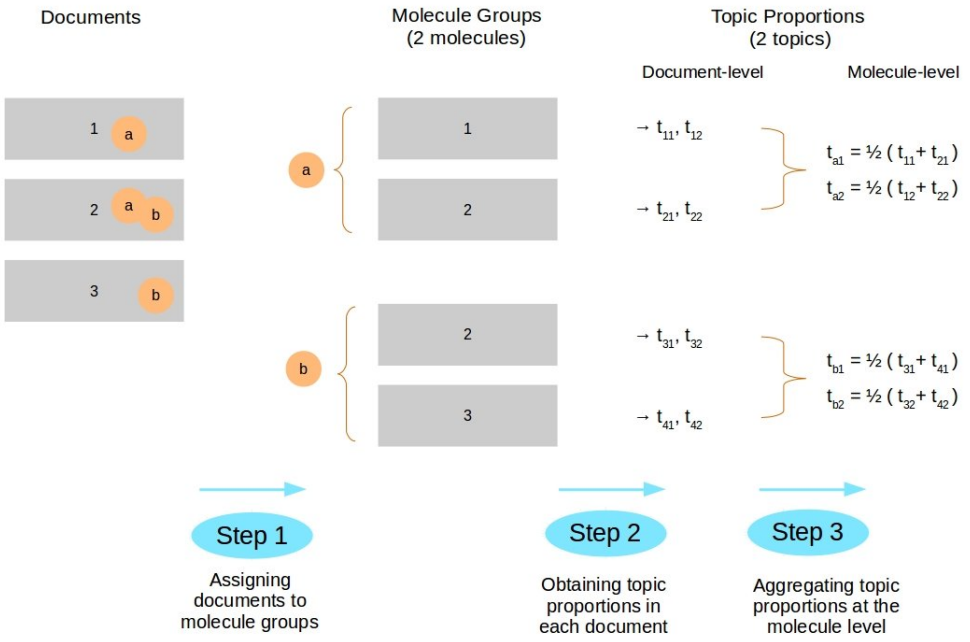
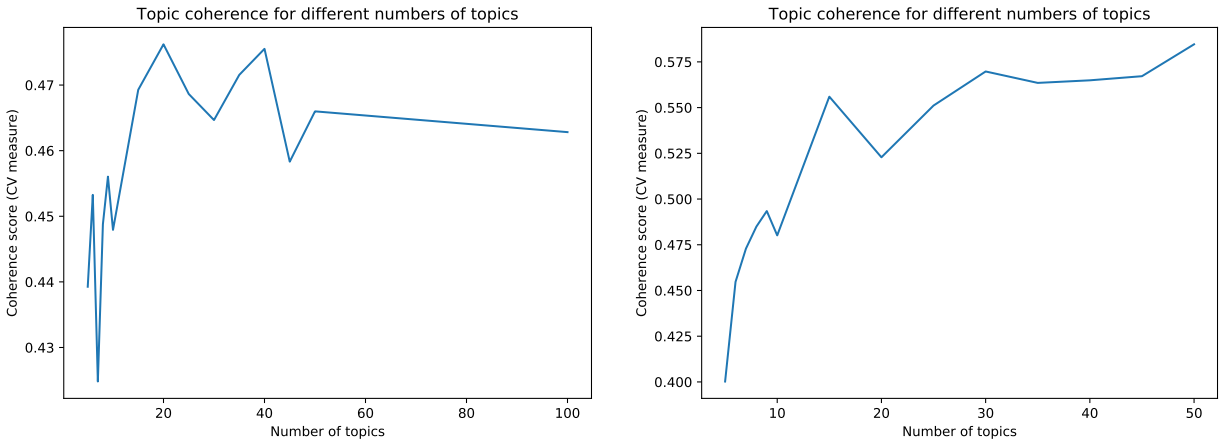


Figure B1: Schematic Explanation of the Methodology

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.

4. 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.



(a) Patents: 20 Topics

(b) Articles: 15 Topics

Figure B2: Topic Coherence Scores

Table B1: Top Twenty Words for Topics in Patents

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5		Topic 6		Topic 7	
Words	Prob	Words	Prob	Words	Prob	Words	Prob	Words	Prob	Words	Prob	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		Topic 9		Topic 10		Topic 11		Topic 12		Topic 13		Topic 14	
Words	Prob	Words	Prob	Words	Prob	Words	Prob	Words	Prob	Words	Prob	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		Topic 16		Topic 17		Topic 18		Topic 19		Topic 20	
Words	Prob	Words	Prob	Words	Prob	Words	Prob	Words	Prob	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 B2: Top Twenty Words for Topics in Articles

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
Words	Prob	Words	Prob	Words	Prob	Words	Prob	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		Topic 7		Topic 8		Topic 9		Topic 10	
Words	Prob	Words	Prob	Words	Prob	Words	Prob	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		Topic 12		Topic 13		Topic 14		Topic 15	
Words	Prob	Words	Prob	Words	Prob	Words	Prob	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

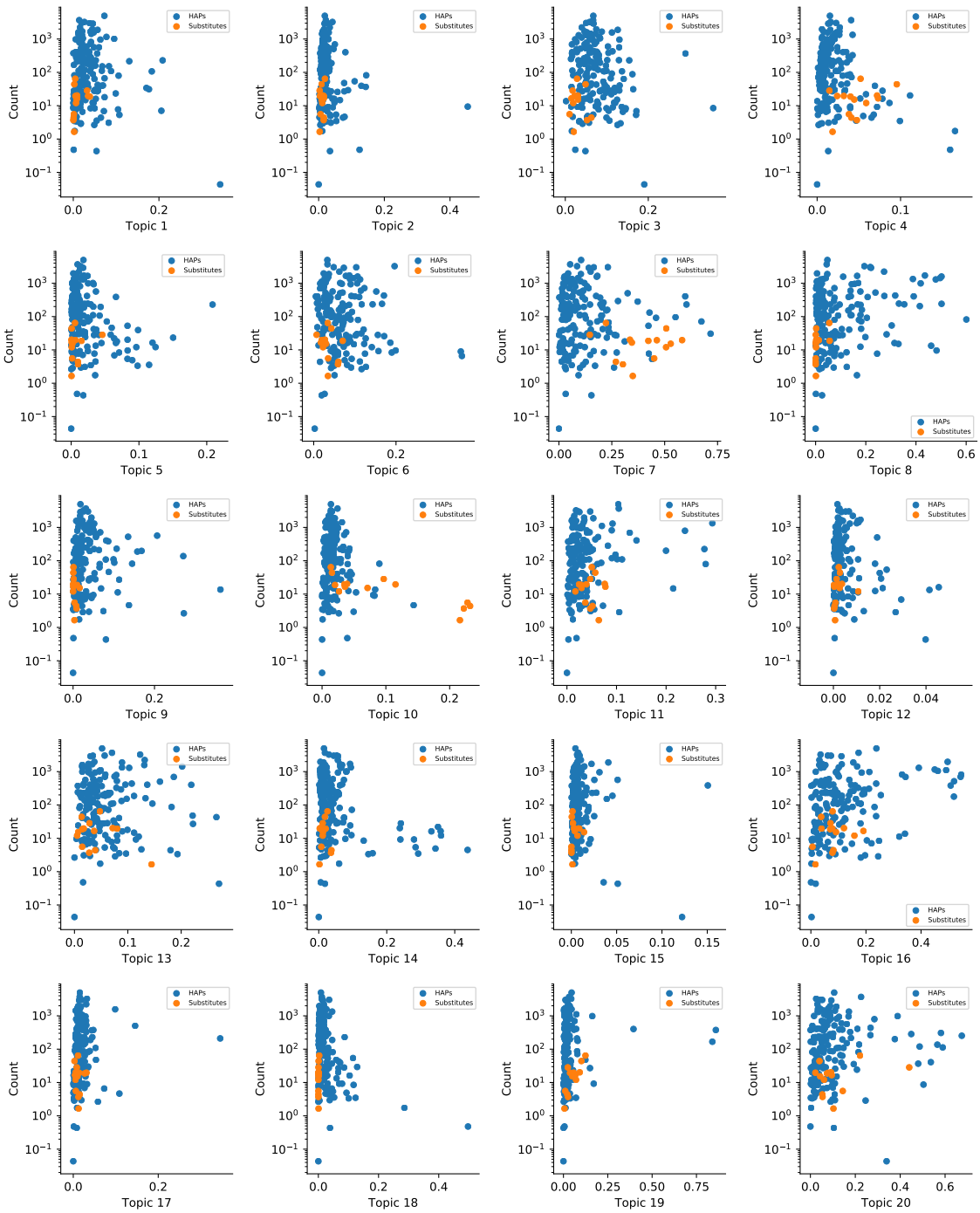


Figure B3: Scatterplot of Topics Proportion and Count for Patents.

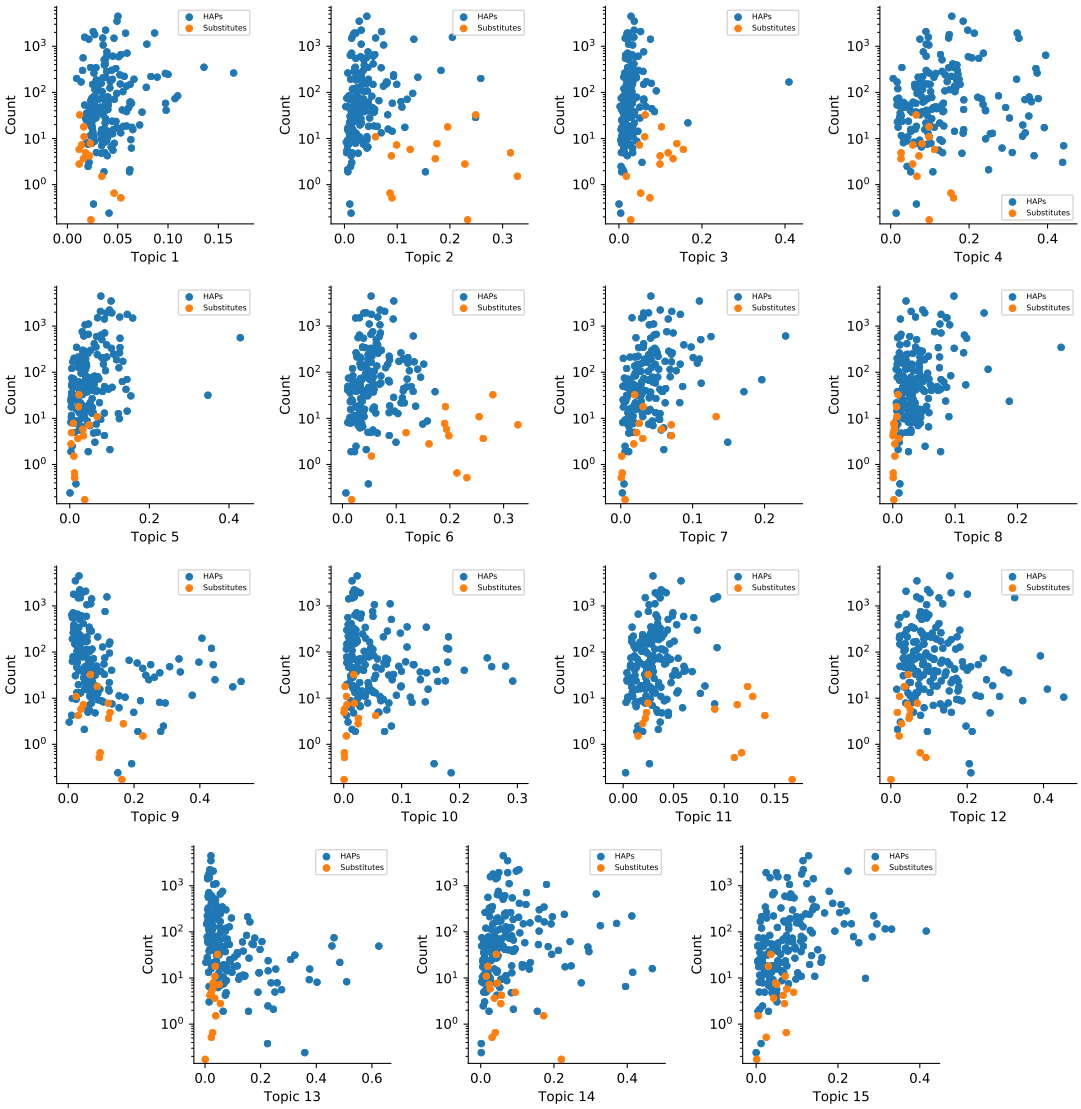
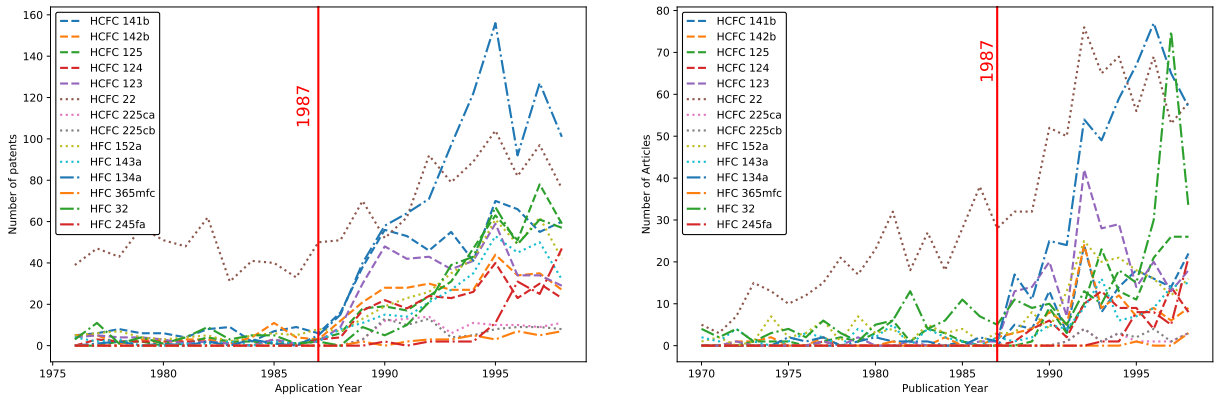


Figure B4: Scatterplot of Topics Proportion and Count for Articles.

C Difference-in-Differences



(a) Patents (b) Articles

Figure C1: Document Counts for Individual CFC Substitutes

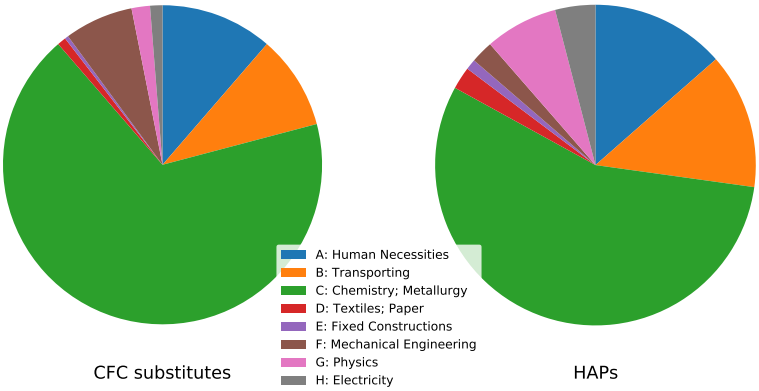


Figure C2: Top Level Patent Codes for CFC Substitutes and HAPs

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 remarkable 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.

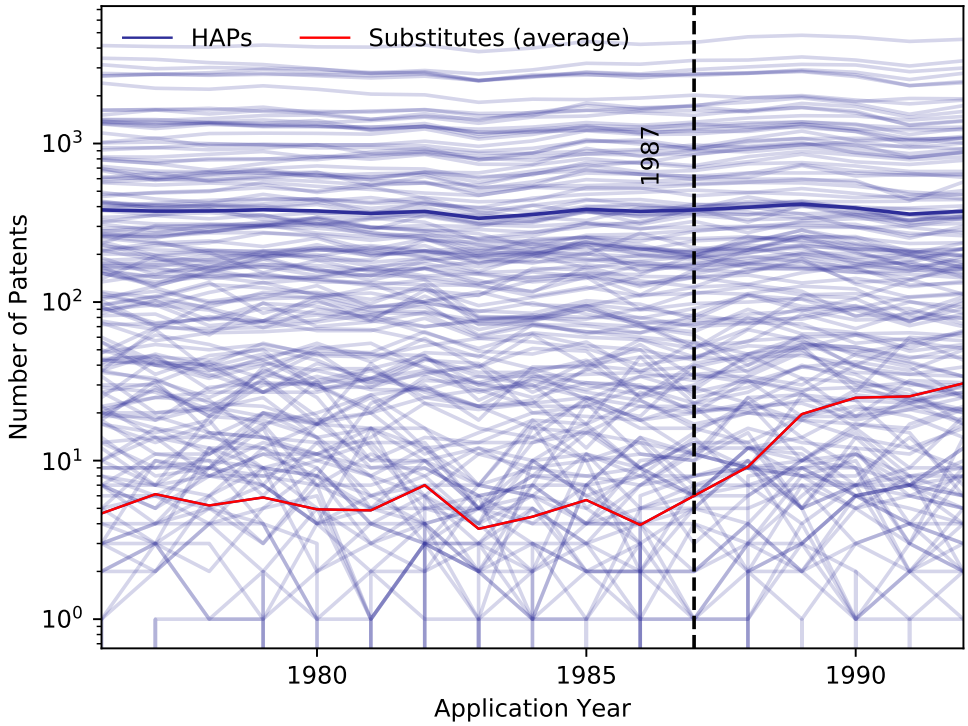


Figure C3: Patent Counts for Each HAP and for the Average CFC Substitute

Note: The graph 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 many HAPs have counts much higher than the average CFC substitute and may, therefore, not be appropriate as comparison units.

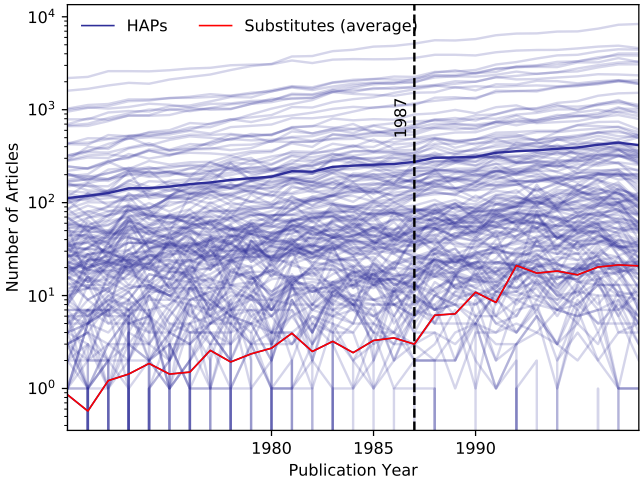


Figure C4: Articles Counts for Each HAP and for the Average CFC Substitute

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 C1: Pre-Period Balance Table Between CFC Substitutes and HAPs

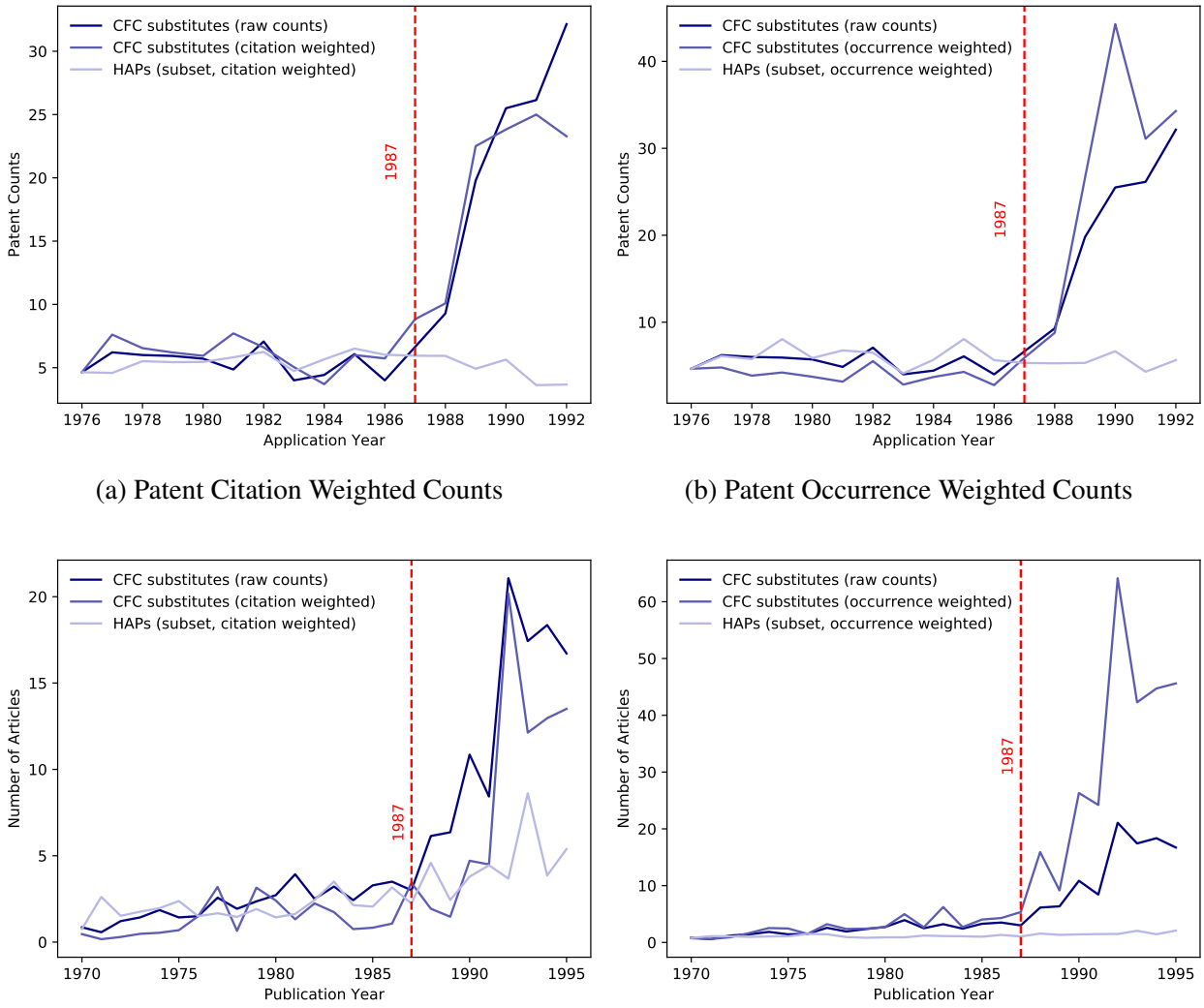
(a) Patents

	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) Patent Citation Weighted Counts

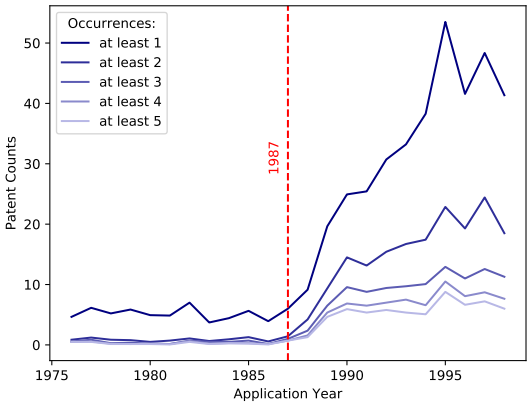
(b) Patent Occurrence Weighted Counts

(c) Article Citation Weighted Counts

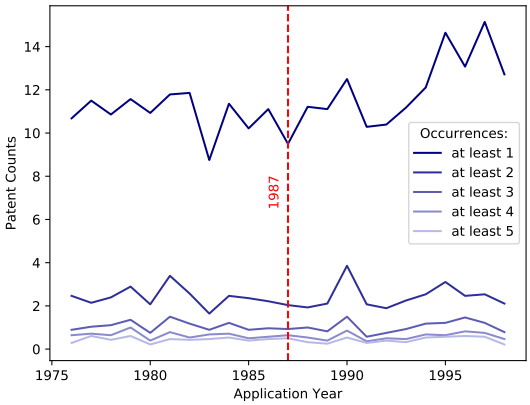
(d) Article Occurrence Weighted Counts

Figure C5: Time Series of Citation- and Occurrence-Weighted Counts

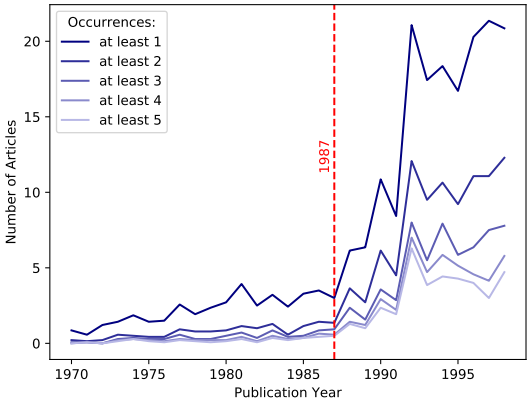
Note: Time-series are scaled to make them equal in the first year of the sample. The graphs indicate that the post-1987 gap between CFC substitutes and HAPs persists even when counts are weighted by the number of citations or by the number of times molecules appear in the text.



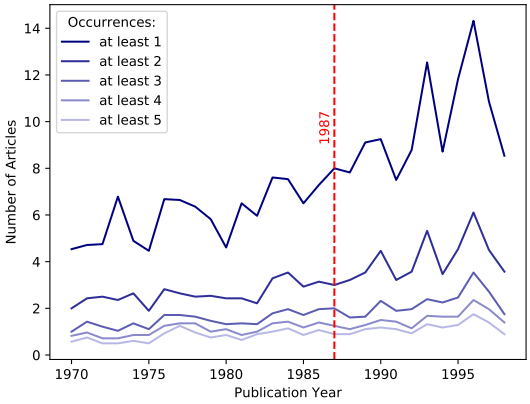
(a) Patents - CFC Substitutes



(b) Patents - HAPs



(c) Articles - CFC Substitutes



(d) Articles - HAPs

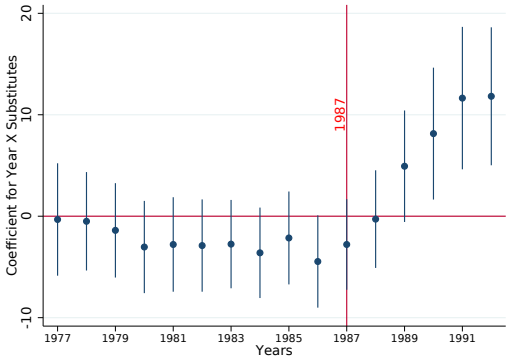
Figure C6: Robustness Check: Counts with Several Thresholds of Molecule Occurrences

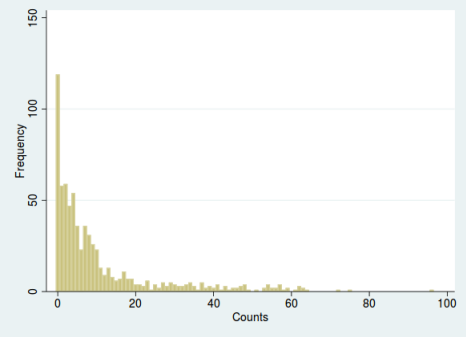
Note: The graphs illustrate that the differential trends CFC substitutes and HAPs are not affected by adopting more stringent definition of what constitutes a document "about CFC substitutes".

Table C2: Difference-in-Differences with Triadic Patents Only

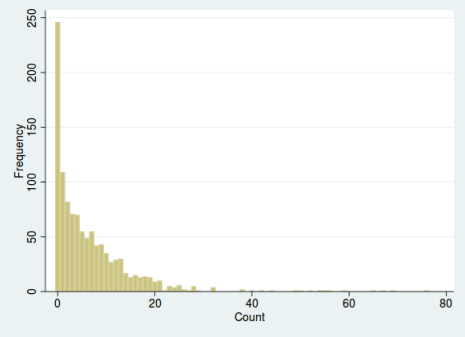
	(1)	(2)
Post 1987 x Substitutes	9.473***	3.370*
	(1.190)	(1.899)
Post 1987 x Substitutes x Years		2.814***
		(0.656)
Substitutes x Years		-0.275**
		(0.121)
Years		0.583***
		(0.071)
Post 1987		-1.555**
		(0.667)
Year FE	Yes	No
Molecule FE	Yes	Yes
R-squared	0.709	0.720
Observations	714	714

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 C7: 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 C3: 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

D Synthetic Control Method

D1 Theoretical Foundations

Here, I briefly summarize the theoretical underpinnings of the synthetic control method. 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, Diamond, and Hainmueller (2010) show that a weighted average of the control units can approximate the counterfactual Y_{it}^N, that is:

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

To understand why this is the case, Equation 1 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 synthetic control method allows the effects of confounding unobserved characteristics to vary with time; taking time differences would not get us rid of μ_i.

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it} \tag{1}$$

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} ... 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.

D2 Figures and Tables

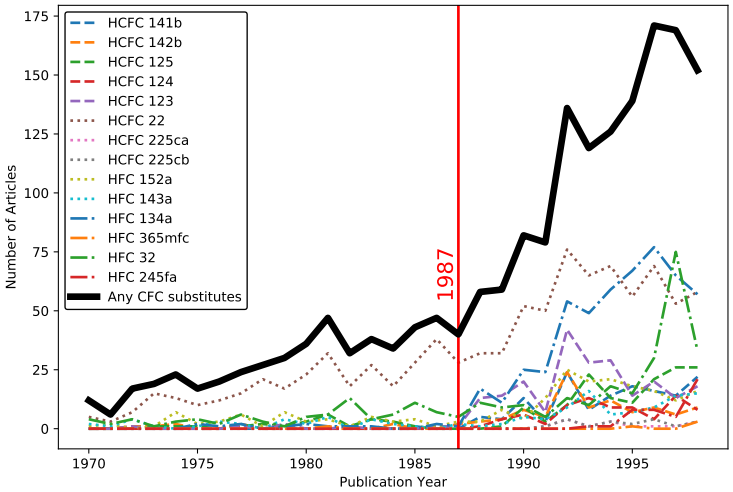


Figure D1: Article Counts for CFC Substitute, Individually and Aggregated

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 D1: Synthetic Control Method 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 synthetic control method algorithm to avoid extrapolation.

Table D2: 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 synthetic control method specification. The information displayed in the "Description" column was collected from the EPA website.

Table D3: Variable Weights Used in the Construction of the Synthetic Control

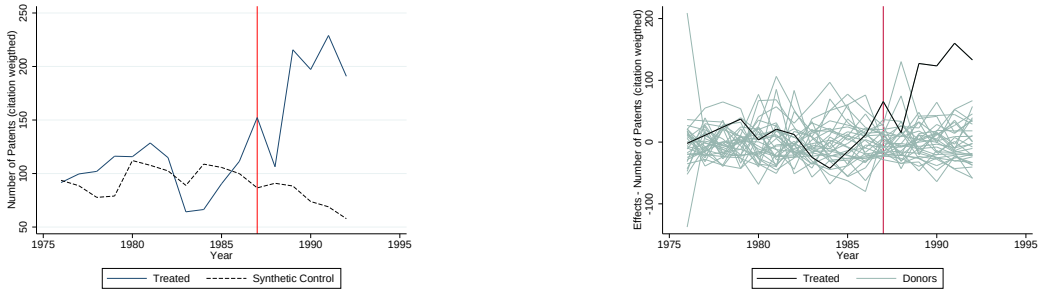
(a) Patents

	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

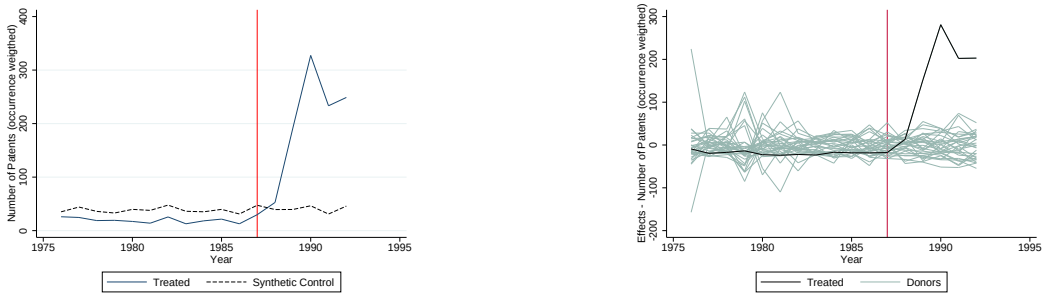
(b) Articles

	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

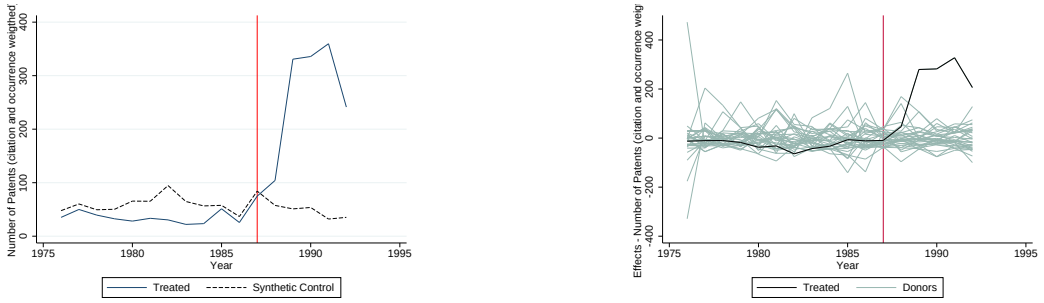
Note: The table displays the value of each variable's contribution to the synthetic control. We note that topic 19, 4 and 7 contribute the most for patents, and topic 11 and 13 for articles. This indicate that these topics had the highest correlations with the outcome variable. In the Stata *synth* package,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



(a) Citation Weighted



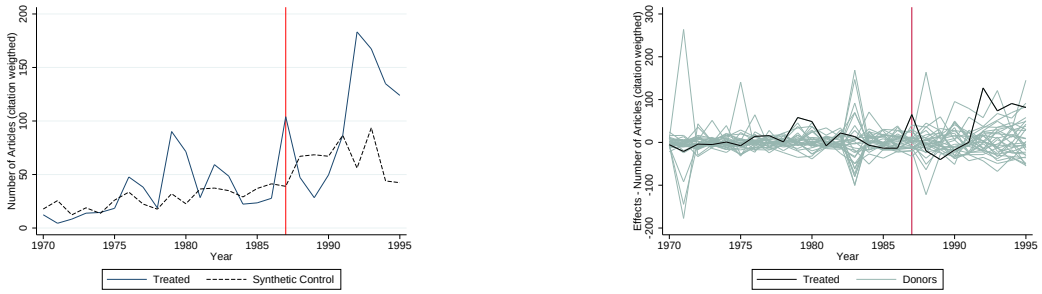
(b) Occurrence Weighted



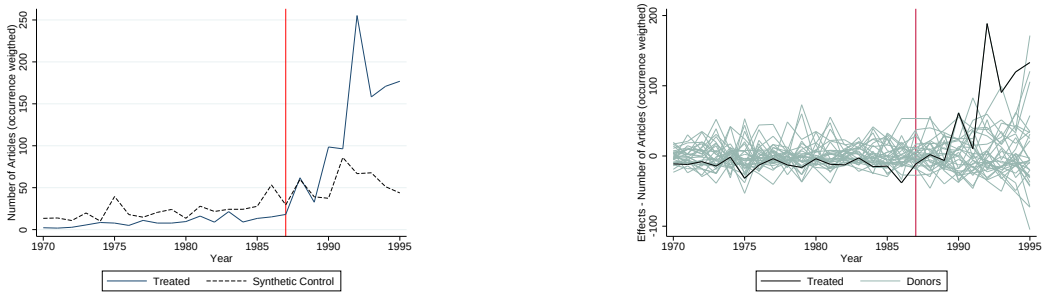
(c) Occurrence and Citation Weighted

Figure D2: Robustness Check for Patents: Synthetic Control Method with Counts Weighted by Occurrences and Citations

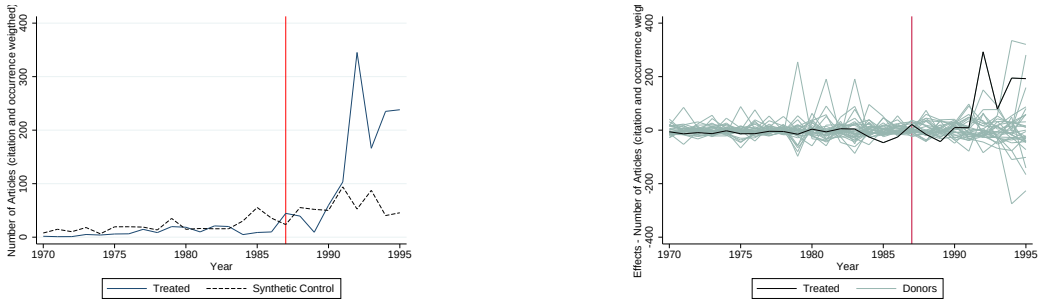
Note: These figures show that implementing the synthetic Control method using patent counts weighted by molecule occurences and patent citation does not alter the main conclusions.



(a) Citation Weighted



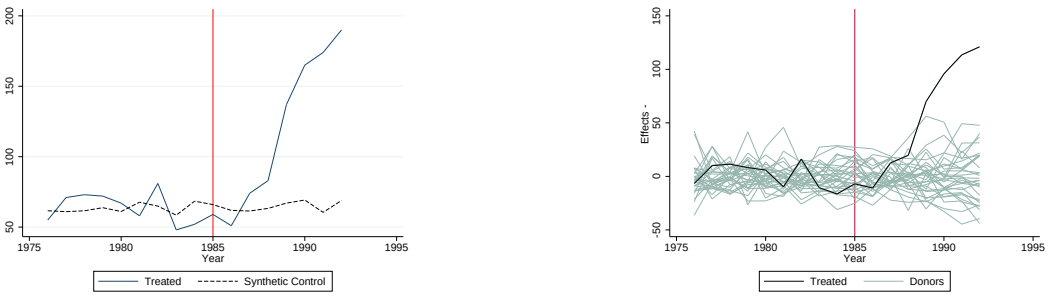
(b) Occurrence Weighted



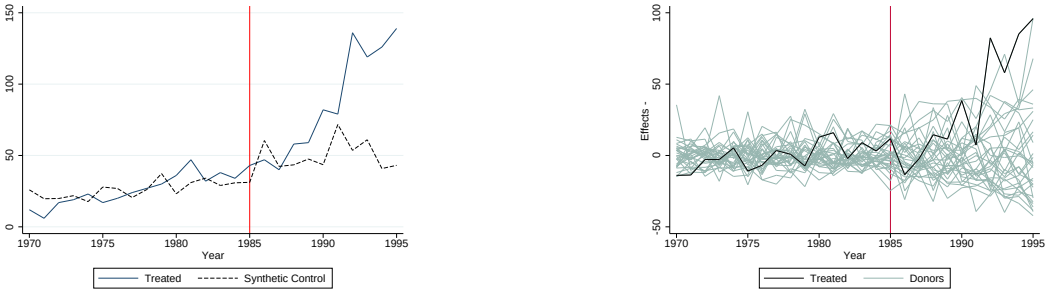
(c) Occurrence and Citation Weighted

Figure D3: Robustness Check for Patents: Synthetic Control Method with Counts Weighted by Occurrences and Citations

Note: These figures show that implementing the synthetic sontrol method 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 D4: Synthetic Control Method Graphs for CFC Substitutes Assuming Anticipation

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

Table E1: 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 with 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 stabilization) of organic compounds. As such, they encompass any patent related to compounds containing carbon and halogen 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 E2: 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 significantly 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 higher 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.

E Others Figures and Tables

Table E3: 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 (e.g., "kinetics" or "evaporative heat transfer") 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.

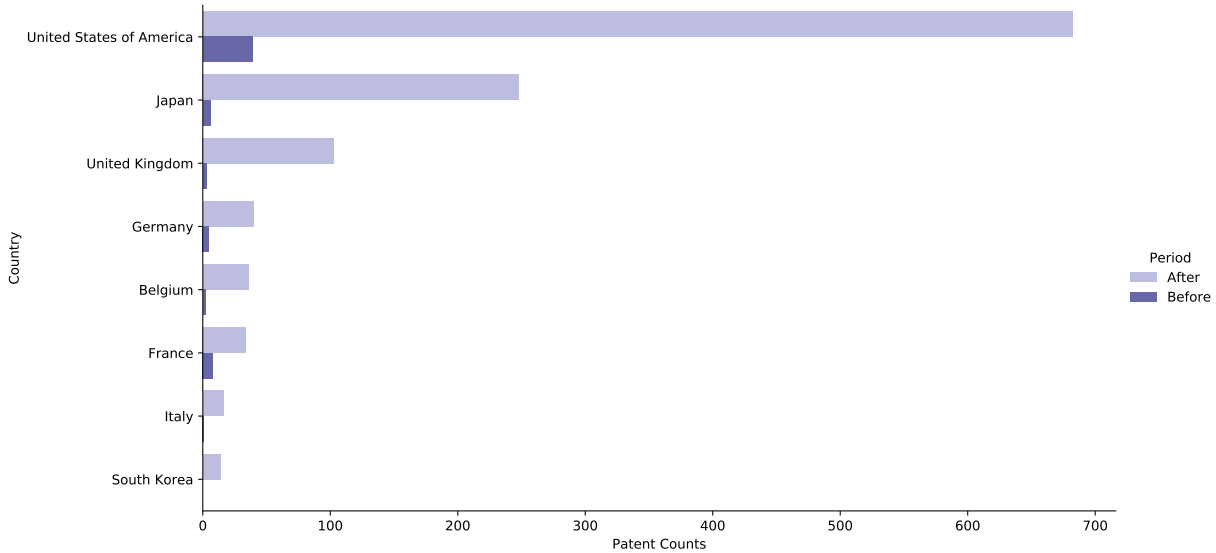


Figure E1: Patent Counts by Country Before and After 1987

Table E4: 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 3 for more details about country and affiliation data.

Table E5: 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 3 for more details about country and affiliation data.

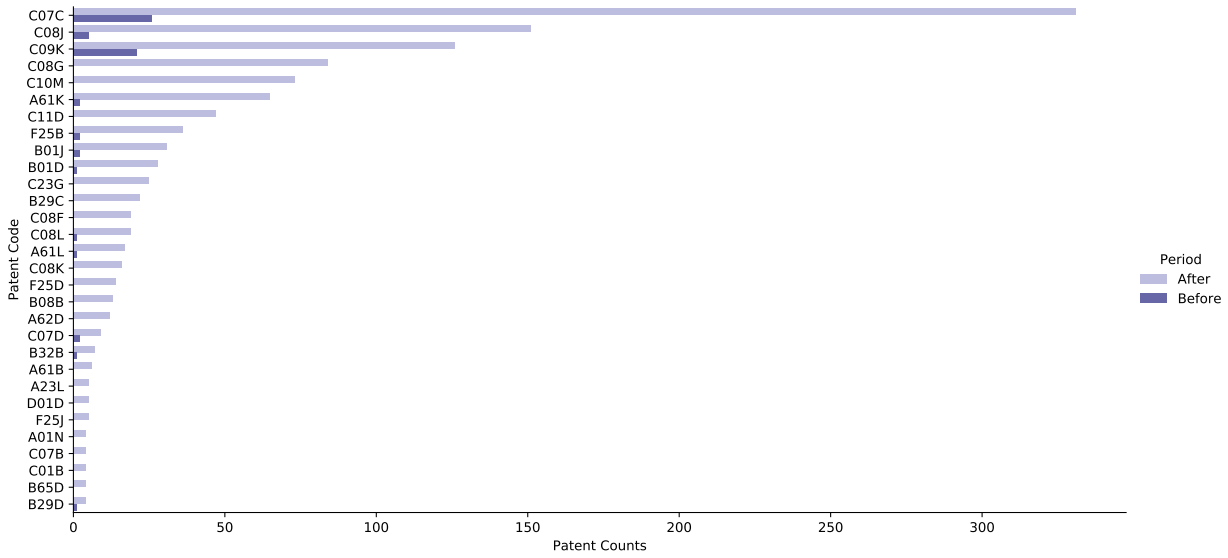


Figure E2: Most Frequent Codes for Patents Mentioning CFC Susbtitutes Before and After 1987

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.

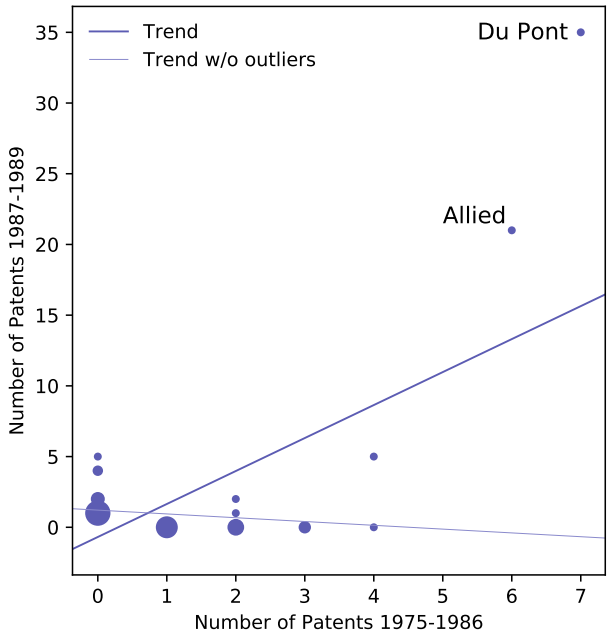


Figure E3: Patenting Before 1987 as a Predictor to Patenting After 1987

Note: The size of the dot is proportional to the number of firms. To limit noise, the sample used to generate the table contains only documents with at least three occurrences of CFC substitutes. The scatter plot shows, for each firm in the sample, patent counts between 1975 and 1986 on the x-axis, and patent counts in the two years that followed Montreal on the y-axis. We see that two outlier firms drive to a positive trend: DuPont and Allied. Excluding those, there are no clear correlations between patenting before 1987 and patenting in the immediate aftermaths of Montreal.

F Theoretical Model

Suppose N countries, all identical and indexed by i. Each country emits a pollutant that damages a shared environmental resource but can also abate an amount q_i of pollution. The benefits from abatement depends on the total amount abated by all countries:

$$B_i(Q) = \frac{b}{N}(aQ - \frac{Q^2}{2}) \tag{2}$$

where $Q = \sum q_i$ and a, b, and c are positive constants.

The costs of abatement only depend on each country's own abatement:

$$C_i(q_i) = \frac{c}{2}q_i^2 \tag{3}$$

At the uncooperative equilibrium, countries abate up to the point where the marignal costs equal the marginal benefits for country i. Hence, we obtain the expression below for q_N, the amount country i abates in the noncooperative equilirbium:

$$MC_i = MB_i \Leftrightarrow cq_i = \frac{b}{N}(a-Q) \Leftrightarrow q_N = \frac{1}{N}\frac{a}{1+\frac{c}{b}} \tag{4}$$

At the cooperative, countries abate up to the point where the marignal costs equal the global marginal benefits. Hence, we obtain the expression below for q_C, the amount country i abates in the cooperative equilirbium:

$$MC_i = \sum_i MB \Leftrightarrow cq_i = N * \frac{b}{N}(a-Q) \Leftrightarrow q_C = \frac{a}{N+\frac{c}{b}} \tag{5}$$

Define the net benefits Π as the difference benefits and costs. The gains from cooeperation are:

$$CooperationGains = \Pi_C - \Pi_N = N * \Big(B_i(q_C) - C_i(q_C)\Big) - N * \Big(B_i(q_N) - C_i(q_N)\Big) \tag{6}$$

Figure F1 illustrates the size of cooperation gains for specific value of b and c (and N set at 100). We note that cooperation gains are highest when c and b are both large. As Barrett (1994) showed, the area when cooperation gains are the highest are is the area where it is the most difficult to sustain a self-enforcing coalition.

Next, I extend this simple model by assuming that countries make their abatement decisions over several time periods and endogenize innovation. The parameter c now is replaced by a function c of the amount of abatement in the previous period:

$$c_t(q_t) = c_0(1-r)^{q_{t-1}} \tag{7}$$

, where c is a constant controlling how costly abatement is, and r a constant between 0 and 1 that can be interpreted as a learning rate. The higher the amount of abatement in period $t-1$ and the lower the marginal cost of abatement in the next period. As Figure F2 illustrates, over several time periods, the area of high gain from cooperation reduces indicating that allocations that used to be difficult to achieve are now within reach.

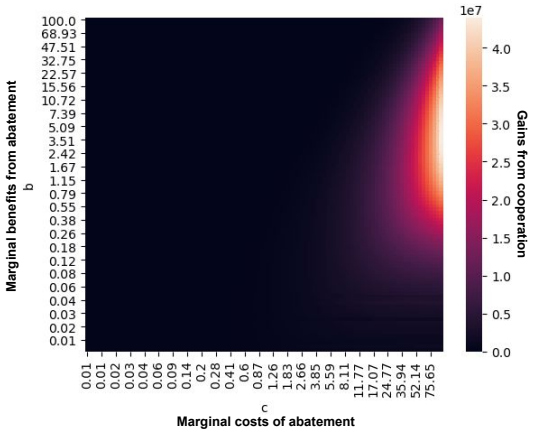
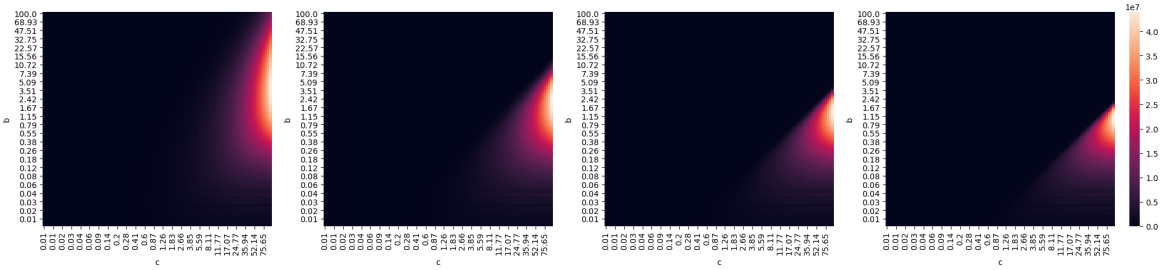


Figure F1: Gains from Cooperation



(a) Period 1 (b) Period 2 (c) Period 3 (d) Period 4

Figure F2: Gains from Cooperation and Induced Innovation