International Environmental Agreements and Directed Technological Change: Evidence from the Ozone Regime
Supporting Online Material

Eugenie Dugoua*

July 2, 2018

*Assistant Professor in Environmental Economics, London School of Economics (starting Sept.). E-mail: eugenie.dugoua@columbia.edu.

Contents

1 Background on Ozone and the Montreal Protocol

- Figure 1 features a timeline of some key events related to CFCs and the Montreal protocol
- Figure 2 displays the actual and projected atmospheric concentrations of several ODS.
- Figure 3 shows the major CFCs by production volume, ozone-depleting potential, industrial sectors and countries by consumption volume in 1986, the year before Montreal was agreed.
- Table 1 shows the name and additional information about the molecules considered under PAFT and AFEAS.
- Table 2 shows the Montreal Protocol phaseout schedules

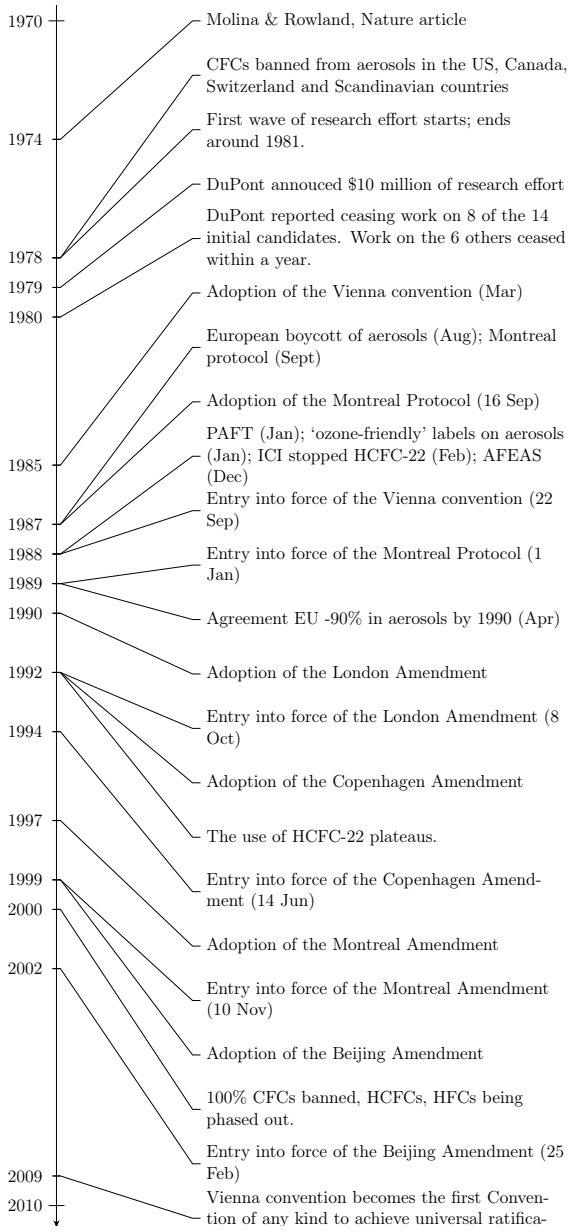


Figure 1: Timeline of events related to CFCs and the Montreal protocol

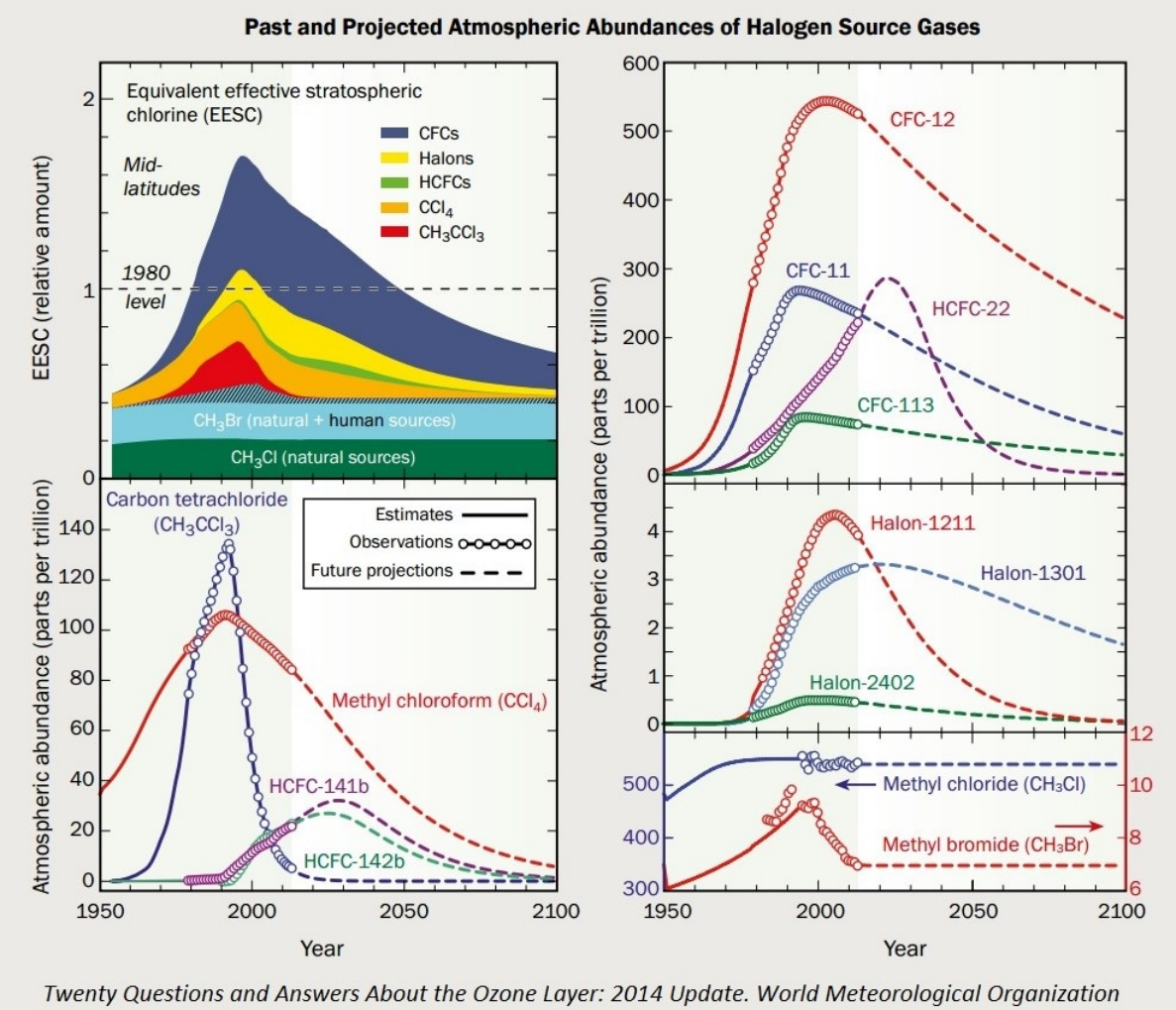


Figure 2: CFC concentrations: past and projected.

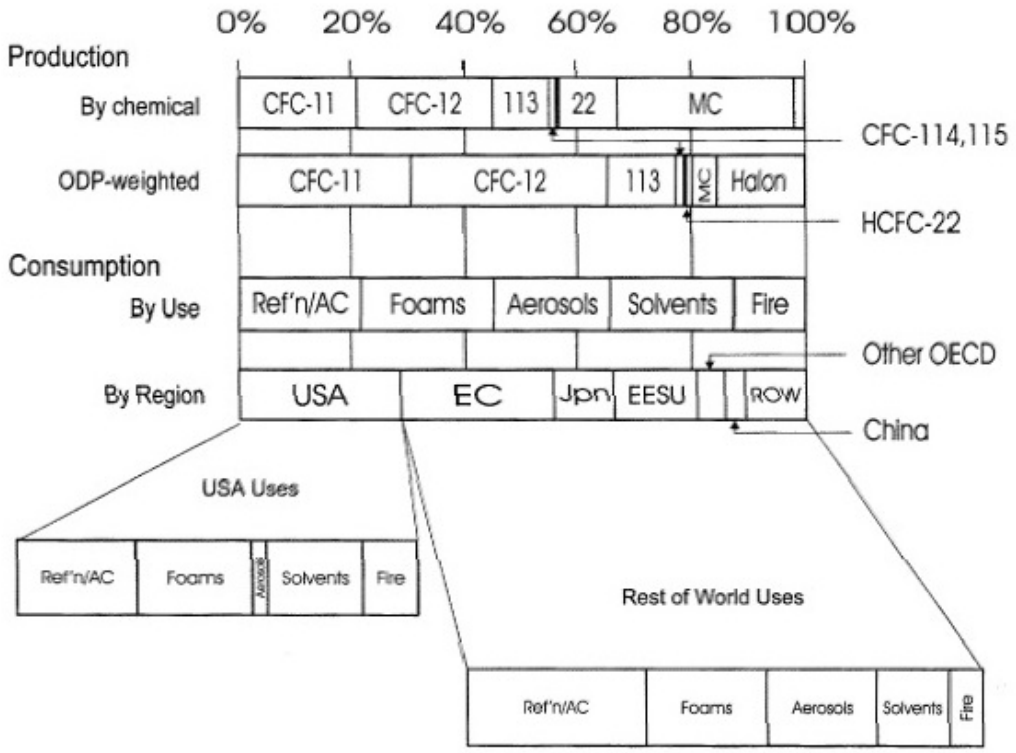


Figure 3: World consumption and production of ozone-depleting chemicals in 1986.
Source: Parson (2003) page 174. Notes: MC stands for methyl chloroform.

Table 1: Details about substitutes to ODS. Information collected from (Parson 2003) and (Benedick 2009). Note: the cost of CFC-12 in 1986 was $0.65/lb.

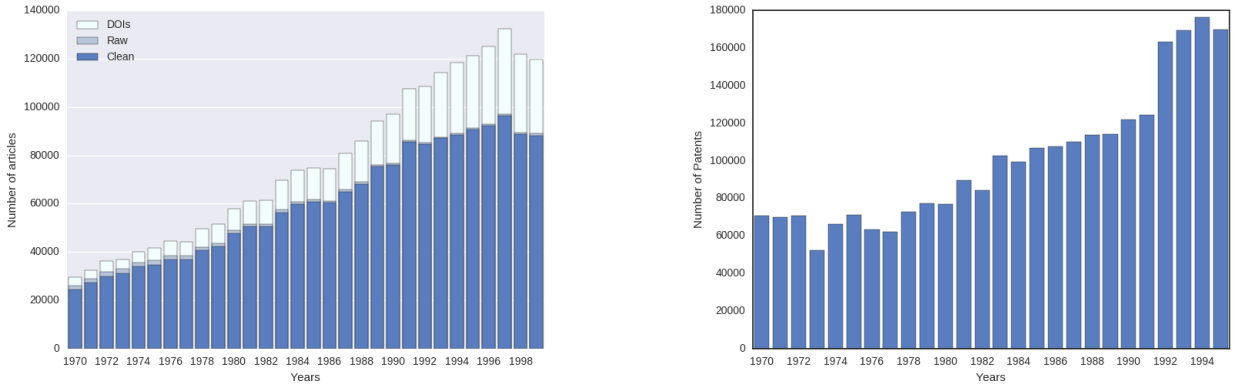
Substitute	PAFT	AFEAS	Substitute for	Notes
HCFC-22	No, already marketed, toxicology known	Yes	Included in Annex C. CFC-11, CFC-12 in foams	cheapest, fastest substitute, already at large scale production at the end of 1986 but due to toxicity concerns, not appropriate for aerosol use. FDA approved it 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	

Table 2: Montreal Protocol Phaseout Schedules. Source: Benedick (2009)

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

2 Data

- Figure 4 displays the number of patents and articles per year
- Table 3 displays the list of molecules for each group: CFC substitutes, HAPs, as well as Annex A and Annex B compounds.
- Table 5 displays the list of substitutes and all their possible names scraped through SciFinder.
- Figure 6 plots the number of documents containing molecules of each group.
- Figure 7 plots the number of documents mentioning Annex A compounds.
- Figure 8 plots the number of documents mentioning Annex B compounds.
- Table 4 displays summary statistics of meta-data for patents for each group of molecules.
- Table 5 displays summary statistics of meta-data for articles for each group of molecules.



(a) Articles

(b) Patents

Figure 4: Total number of documents collected per year

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
Îś-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)

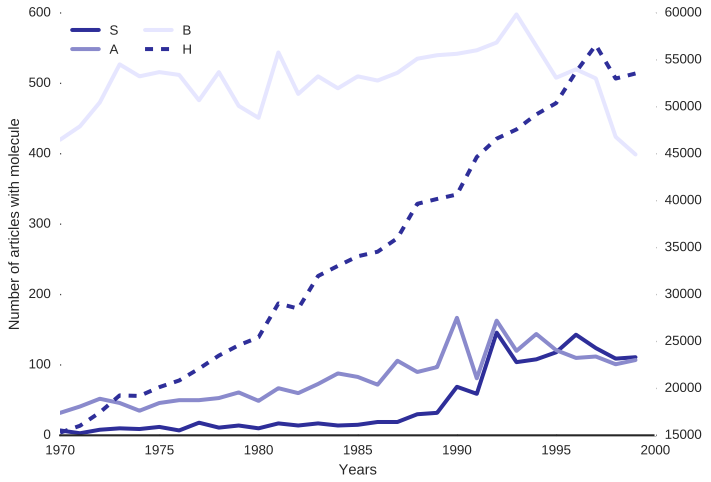
HFC 365mfc
1,1,1,3,3-Pentafluorobutane
2,2,4,4,4-Pentafluorobutane

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

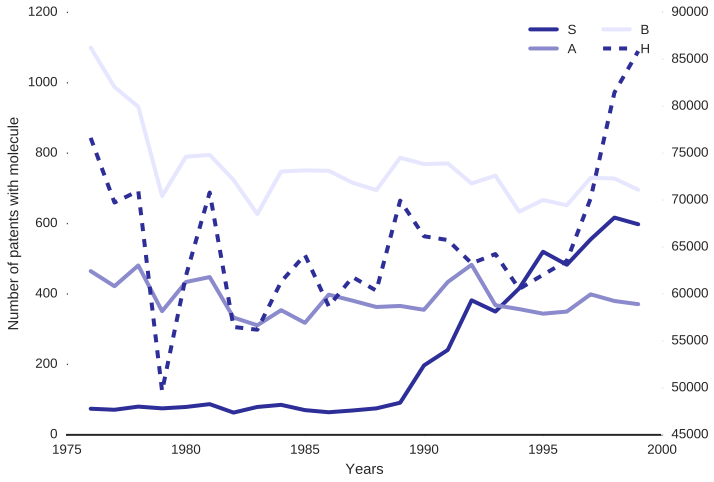
Figure 5: List of substitutes and their possible names

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 3: List molecules in each treatment group

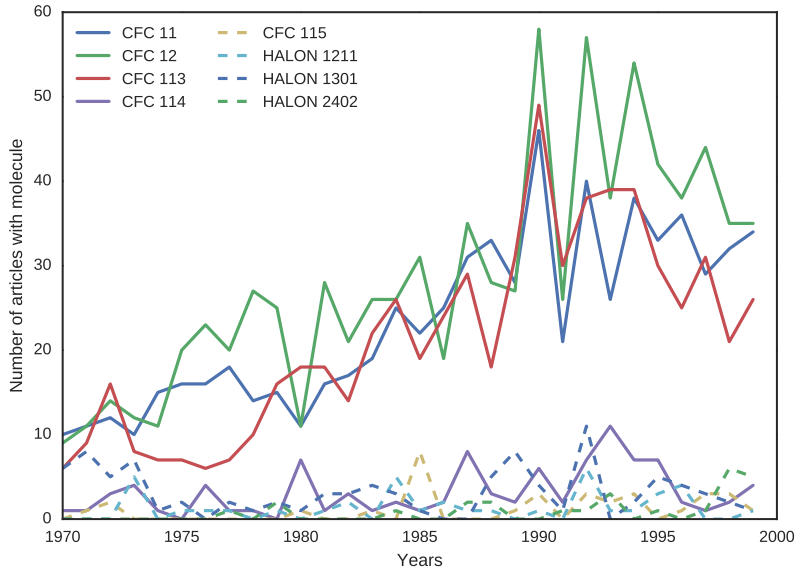


(a) Articles

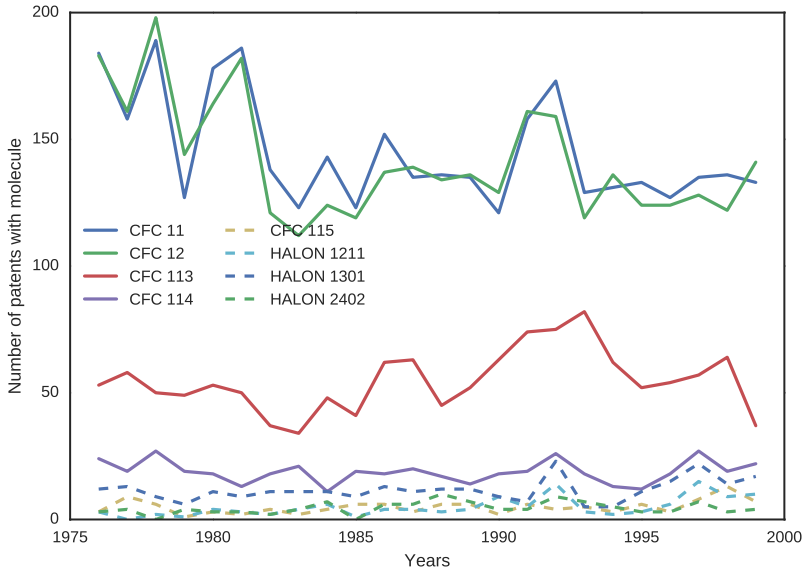


(b) Patents

Figure 6: Total number of documents mentioning CFC substitutes (S), Annex A compounds (A), Annex B compounds (B) and HAPS (H). The right axis is relative to HAPs.

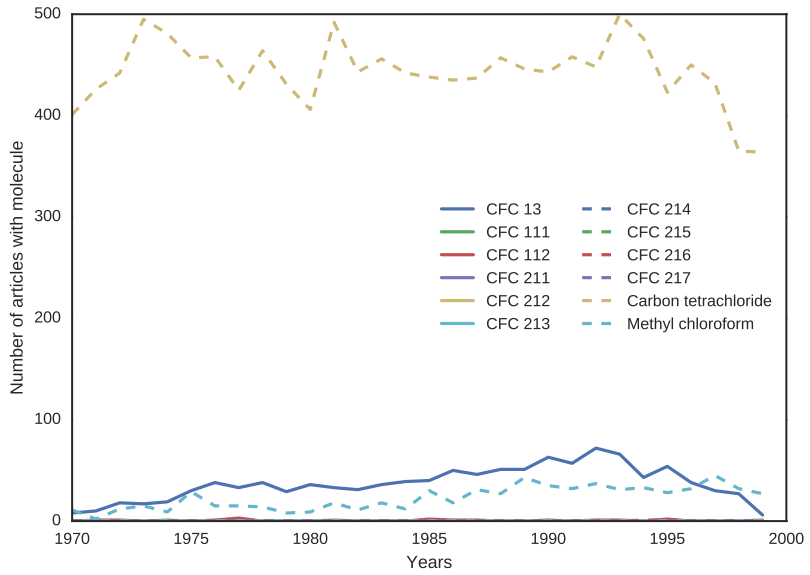


(a) Number of articles containing molecules from group A.

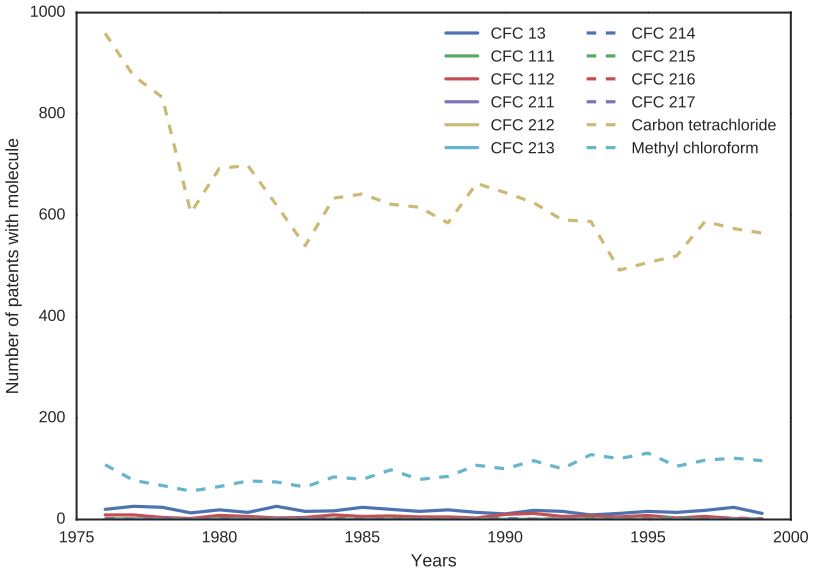


(b) Number of patents containing molecules from group A.

Figure 7: Total number of documents containing names of molecules from groups A.



(a) Number of articles containing molecules from group B.



(b) Number of patents containing molecules from group B.

Figure 8: Total number of documents containing names of molecules from groups B.

	Annex A	Annex B	CFC Substitutes	HAPs
Year	1988.38	1987.90	1992.55	1989.61
	(7.49)	(7.67)	(7.09)	(7.69)
Education	0.02	0.01	0.02	0.02
	(0.12)	(0.11)	(0.14)	(0.14)
Company	0.97	0.97	0.97	0.96
	(0.18)	(0.16)	(0.17)	(0.20)
Government	0.02	0.01	0.01	0.02
	(0.13)	(0.10)	(0.07)	(0.13)
Facility	0.00	0.00	0.00	0.00
	(0.03)	(0.03)	(0.07)	(0.02)
Nonprofit	0.00	0.00	0.00	0.00
	(0.04)	(0.05)	(0.00)	(0.07)
Healthcare	0.00	0.00	0.00	0.00
	(0.00)	(0.01)	(0.00)	(0.02)
USA	0.59	0.47	0.61	0.56
	(0.49)	(0.50)	(0.49)	(0.50)
Europe	0.27	0.29	0.21	0.23
	(0.44)	(0.45)	(0.40)	(0.42)
Japan	0.12	0.23	0.17	0.17
	(0.32)	(0.42)	(0.37)	(0.38)

Table 4: Summary statistics of meta-data for patents per group

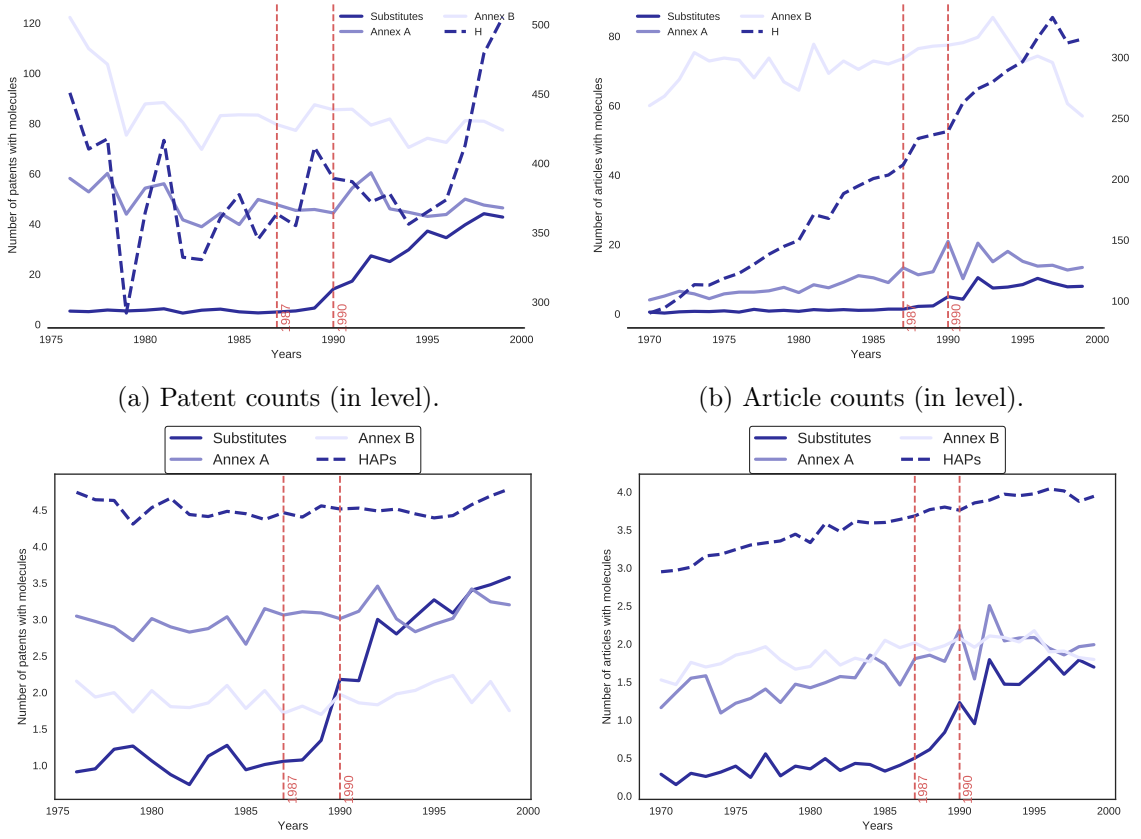
	Annex A	Annex B	CFC Substitutes
Year	1986.88	1984.11	1989.78
	(7.52)	(8.17)	(7.25)
Citation Count	25.29	25.42	30.60
	(55.75)	(147.84)	(72.19)
Number of Authors	3.35	2.98	2.95
	(5.33)	(2.09)	(2.85)
Education	0.71	0.79	0.75
	(0.45)	(0.41)	(0.43)
Company	0.12	0.07	0.12
	(0.33)	(0.25)	(0.32)
Government	0.08	0.08	0.11
	(0.28)	(0.27)	(0.31)
Facility	0.16	0.11	0.11
	(0.36)	(0.31)	(0.31)
Nonprofit	0.02	0.02	0.02
	(0.15)	(0.13)	(0.13)
Healthcare	0.02	0.02	0.02
	(0.14)	(0.15)	(0.15)
USA	0.41	0.27	0.36
	(0.49)	(0.44)	(0.48)
Europe	0.39	0.45	0.39
	(0.49)	(0.50)	(0.49)
Japan	0.09	0.10	0.12
	(0.29)	(0.31)	(0.32)

Table 5: Meta-Data of Articles per Group

3 Difference-in-Differences

- Figure 9 displays the average counts and log counts for the different molecule groups using the weak rule.
- Figure 10 shows the mean counts in articles for the different groups with method weak, intermediate and strong.
- Figure 11 shows the mean counts in patents for the different groups with method weak, intermediate and strong.
- Table 8 displays DiD specifications with bootstrap for patents for CFC substitutes.
- Table 9 displays DiD specifications with bootstrap for articles for CFC substitutes.
- Figure 12 displays DiD specifications with treatment effects by year for patents for CFC substitutes.
- Figure 13 displays DiD specifications with treatment effects by year for articles for CFC substitutes.
- Table 8 displays DiD specifications controlling for topic proportions for patents for CFC substitutes.
- Table 9 displays DiD specifications controlling for topic proportions for articles for CFC substitutes.
- Table 12 displays DiD specifications with bootstrap for patents for Annex A compounds.
- Table 13 displays DiD specifications with bootstrap for articles for Annex A compounds.
- Figure 15 displays DiD specifications with treatment effects by year for patents for Annex A compounds.
- Figure 14 displays DiD specifications with treatment effects by year for articles for Annex A compounds.
- Table 12 displays DiD specifications controlling for topic proportions for patents for Annex A compounds.
- Table 13 displays DiD specifications controlling for topic proportions for articles for Annex A compounds.
- Table 16 displays DiD specifications with bootstrap for patents for Annex B compounds.

- Table 17 displays DiD specifications with bootstrap for articles for Annex B compounds.
- Figure 17 displays DiD specifications with treatment effects by year for patents for Annex B compounds.
- Figure 16 displays DiD specifications with treatment effects by year for articles for Annex B compounds.
- Table 16 displays DiD specifications controlling for topic proportions for patents for Annex B compounds.
- Table 17 displays DiD specifications controlling for topic proportions for articles for Annex B compounds.



(a) Patent counts (in level).

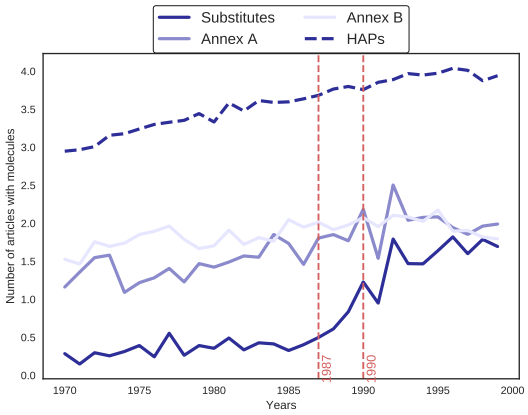
(b) Article counts (in level).

(c) Patent counts (in log).

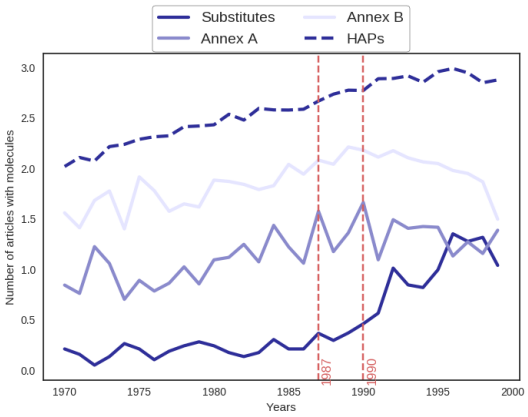
(d) Article counts (in log).

Figure 9: Mean Document Counts and Log Counts for the Different Molecule Groups Using the Weak Rule.

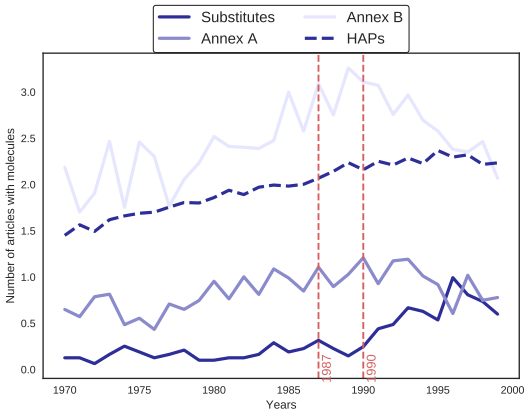
Notes: Here, I plot the yearly mean counts of documents related to each of the four molecule groups to gauge similarity in the pre-trends. In the article corpus, HAPs and Annex A seem to have a clear upward trend before 1987 while substitutes and Annex B seem somewhat flat. Since pre-trends appear to differ, all of the HAP molecules can not serve as good control.



(a) Weak.

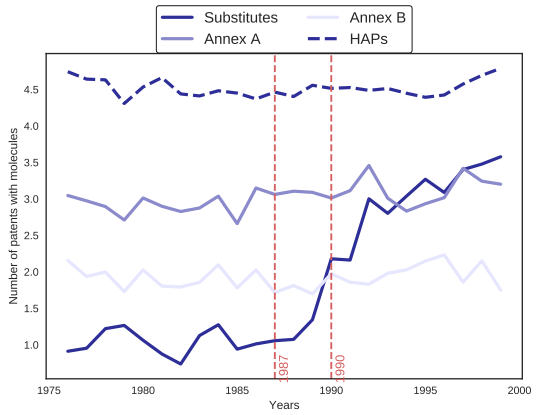


(b) Intermediate.

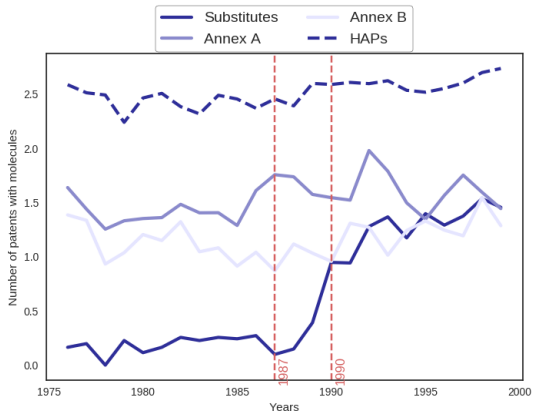


(c) Strong.

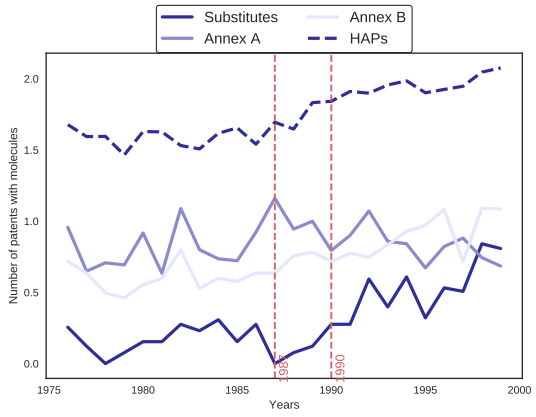
Figure 10: Mean counts in log for the different groups for articles.



(a) Weak.



(b) Intermediate.



(c) Strong.

Figure 11: Mean counts in log for the different groups for articles.

Table 6: Patents: Diff-in-Diff. Additional Specifications

	(1)	(2)	(3)	(4)
Post 1987 x Substitutes	1.637***	1.637***		
	(0.080)	(0.078)		
Post 1987 x Substitutes x Years			0.215***	0.215***
			(0.018)	(0.020)
Substitutes x Years			0.014	0.014
			(0.011)	(0.012)
Years			0.003**	0.003**
			(0.001)	(0.001)
Year FE	Yes	Yes	No	No
Molecule FE	Yes	Yes	Yes	Yes
Bootstraped	No	Yes	No	Yes
R-squared	0.952	0.952	0.962	0.962
Observations	1344	1344	1344	1344

Standard errors in parentheses
Dependent variable: Log count of patents
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Articles: Diff-in-Diff. Additional Specifications

	(1)	(2)	(3)	(4)
Post 1987 x Substitutes	0.668***	0.668***		
	(0.068)	(0.071)		
Post 1987 x Substitutes x Years			0.100***	0.100***
			(0.013)	(0.013)
Substitutes x Years			-0.003	-0.003
			(0.006)	(0.006)
Years			0.024***	0.024***
			(0.001)	(0.002)
Year FE	Yes	Yes	No	No
Molecule FE	Yes	Yes	Yes	Yes
Bootstraped	No	Yes	No	Yes
R-squared	0.951	0.951	0.950	0.950
Observations	1680	1680	1680	1680

Standard errors in parentheses
Dependent variable: Log count of articles
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

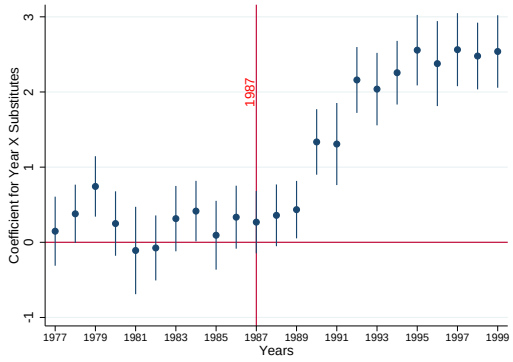
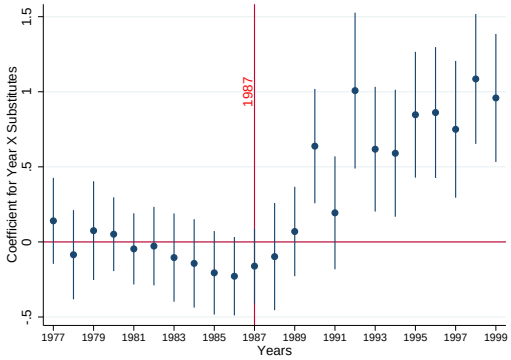


Figure 12: Patents: Differences-in-difference treatment effects by year

Figure 13: Articles: Differences-in-difference treatment effects by year



	(1)	(2)	(3)	(4)	(5)	(6)
Post 1987 x Substitutes	1.637***	1.298***	1.253***			
	(0.080)	(0.079)	(0.080)			
Weighted mean proportion of topic 1		2.197***			1.446**	
		(0.689)			(0.615)	
Weighted mean proportion of topic 2		0.330			0.130	
		(0.768)			(0.607)	
Weighted mean proportion of topic 3		1.685***			1.047***	
		(0.396)			(0.347)	
Weighted mean proportion of topic 4		1.072			1.613**	
		(0.674)			(0.645)	
Weighted mean proportion of topic 5		-2.633***			-1.332**	
		(0.808)			(0.655)	
Mean proportion of topic 1			3.491***			1.978**
			(0.932)			(0.820)
Mean proportion of topic 2			0.256			0.221
			(0.831)			(0.676)
Mean proportion of topic 3			1.748***			1.514***
			(0.499)			(0.414)
Mean proportion of topic 4			0.616			0.829
			(0.854)			(0.708)
Mean proportion of topic 5			-4.380***			-2.908***
			(1.069)			(0.825)
Post 1987 x Substitutes x Years				0.215***	0.193***	0.188***
				(0.018)	(0.017)	(0.017)
Substitutes x Years				0.014	0.004	0.005
				(0.011)	(0.009)	(0.009)
Years				0.003**	0.003**	0.003**
				(0.001)	(0.001)	(0.001)
Year FE	Yes	Yes	Yes	No	No	No
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.952	0.961	0.962	0.962	0.968	0.968
Observations	1344	1344	1344	1344	1344	1344

Standard errors in parentheses
Dependent variable: Log count of patents
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Patents: Diff-in-Diff. Controlling for topic proportions.

	(1)	(2)	(3)	(4)	(5)	(6)
Post 1987 x Substitutes	0.668***	0.224***	0.214***			
	(0.068)	(0.057)	(0.058)			
Weighted mean proportion of topic 1		0.905**			1.058***	
		(0.403)			(0.390)	
Weighted mean proportion of topic 2		1.008***			0.936***	
		(0.301)			(0.298)	
Weighted mean proportion of topic 3		1.432***			1.140**	
		(0.448)			(0.454)	
Weighted mean proportion of topic 4		1.211***			1.234***	
		(0.357)			(0.347)	
Weighted mean proportion of topic 5		1.077***			1.113***	
		(0.316)			(0.320)	
Mean proportion of topic 1			1.484***			1.659***
			(0.478)			(0.466)
Mean proportion of topic 2			0.594			0.448
			(0.378)			(0.370)
Mean proportion of topic 3			1.415***			1.122**
			(0.533)			(0.541)
Mean proportion of topic 4			1.360***			1.446***
			(0.460)			(0.451)
Mean proportion of topic 5			0.991**			1.052***
			(0.390)			(0.393)
Post 1987 x Substitutes x Years				0.100***	0.051***	0.051***
				(0.013)	(0.011)	(0.011)
Substitutes x Years				-0.003	-0.008*	-0.009**
				(0.006)	(0.004)	(0.004)
Years				0.024***	0.022***	0.022***
				(0.001)	(0.001)	(0.001)
Year FE	Yes	Yes	Yes	No	No	No
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.951	0.964	0.964	0.950	0.963	0.964
Observations	1680	1680	1680	1680	1680	1680

Standard errors in parentheses
Dependent variable: Log count of articles
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Articles: Diff-in-Diff. Controlling for topic proportions.

	(1)	(2)	(3)	(4)
Post 1987 x Annex A	0.168***	0.168***		
	(0.058)	(0.056)		
Post 1987 x Annex A x Years			0.007	0.007
			(0.016)	(0.015)
Annex A x Years			0.008	0.008
			(0.010)	(0.010)
Years			0.002	0.002
			(0.002)	(0.002)
Year FE	Yes	Yes	No	No
Molecule FE	Yes	Yes	Yes	Yes
Bootstraped	No	Yes	No	Yes
R-squared	0.972	0.972	0.968	0.968
Observations	768	768	768	768

Standard errors in parentheses
Dependent variable: Log count of patents
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Patents: Diff-in-Diff. Additional Specifications

	(1)	(2)	(3)	(4)
Post 1987 x Annex A	0.069	0.069		
	(0.073)	(0.080)		
Post 1987 x Annex A x Years			-0.008	-0.008
			(0.017)	(0.019)
Annex A x Years			0.006	0.006
			(0.008)	(0.008)
Years			0.028***	0.028***
			(0.002)	(0.002)
Year FE	Yes	Yes	No	No
Molecule FE	Yes	Yes	Yes	Yes
Bootstraped	No	Yes	No	Yes
R-squared	0.966	0.966	0.964	0.964
Observations	960	960	960	960

Standard errors in parentheses
Dependent variable: Log count of articles
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Articles: Diff-in-Diff. Additional Specifications

Figure 14: Patents: Differences-in-difference treatment effects by year

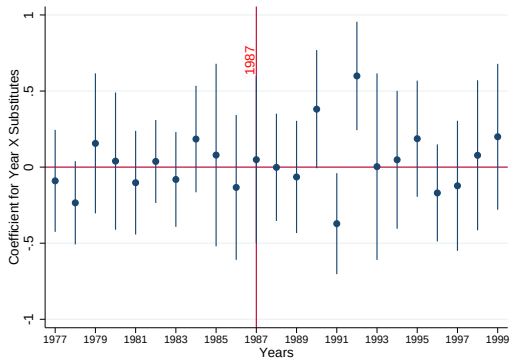


Figure 15: Articles: Differences-in-difference treatment effects by year

	(1)	(2)	(3)	(4)	(5)	(6)
Post 1987 x Annex A	0.168***	0.127**	0.127**			
	(0.058)	(0.050)	(0.051)			
Weighted mean proportion of topic 1		2.057***			1.645**	
		(0.770)			(0.802)	
Weighted mean proportion of topic 2		0.956			1.025	
		(0.776)			(0.806)	
Weighted mean proportion of topic 3		1.454***			1.406***	
		(0.376)			(0.377)	
Weighted mean proportion of topic 4		1.978***			2.242***	
		(0.599)			(0.624)	
Weighted mean proportion of topic 5		1.075			1.345	
		(0.821)			(0.837)	
Mean proportion of topic 1			2.080**			1.505
			(0.928)			(0.970)
Mean proportion of topic 2			1.171			1.264
			(0.910)			(0.948)
Mean proportion of topic 3			1.408***			1.416***
			(0.435)			(0.431)
Mean proportion of topic 4			2.019***			2.339***
			(0.777)			(0.780)
Mean proportion of topic 5			0.920			1.154
			(1.143)			(1.158)
Post 1987 x Annex A x Years				0.007	0.015	0.013
				(0.016)	(0.013)	(0.013)
Annex A x Years				0.008	0.001	0.002
				(0.010)	(0.008)	(0.008)
Years				0.002	0.002	0.002
				(0.002)	(0.002)	(0.002)
Year FE	Yes	Yes	Yes	No	No	No
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.972	0.977	0.977	0.968	0.973	0.973
Observations	768	768	768	768	768	768

Standard errors in parentheses
Dependent variable: Log count of patents
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Patents: Diff-in-Diff. Controlling for topic proportions.

	(1)	(2)	(3)	(4)	(5)	(6)
Post 1987 x Annex A	0.069	-0.040	-0.052			
	(0.073)	(0.060)	(0.059)			
Weighted mean proportion of topic 1		0.763*			1.111**	
		(0.422)			(0.440)	
Weighted mean proportion of topic 2		0.293			0.286	
		(0.448)			(0.442)	
Weighted mean proportion of topic 3		2.199***			1.973***	
		(0.553)			(0.554)	
Weighted mean proportion of topic 4		1.008**			0.922**	
		(0.448)			(0.449)	
Weighted mean proportion of topic 5		1.522***			1.488***	
		(0.368)			(0.367)	
Mean proportion of topic 1			1.624***			2.039***
			(0.477)			(0.497)
Mean proportion of topic 2			-0.244			-0.275
			(0.554)			(0.553)
Mean proportion of topic 3			2.292***			2.006***
			(0.673)			(0.675)
Mean proportion of topic 4			0.650			0.600
			(0.569)			(0.580)
Mean proportion of topic 5			1.881***			1.809***
			(0.468)			(0.459)
Post 1987 x Annex A x Years				-0.008	-0.015	-0.016
				(0.017)	(0.013)	(0.013)
Annex A x Years				0.006	0.002	0.002
				(0.008)	(0.006)	(0.006)
Years				0.028***	0.027***	0.026***
				(0.002)	(0.002)	(0.002)
Year FE	Yes	Yes	Yes	No	No	No
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.966	0.976	0.976	0.964	0.974	0.975
Observations	960	960	960	960	960	960

Standard errors in parentheses
Dependent variable: Log count of articles
Years are relative to 1987.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Articles: Diff-in-Diff. Controlling for topic proportions.

	(1)	(2)	(3)	(4)
Post 1990 x Annex B	-0.002	-0.002		
	(0.061)	(0.061)		
Post 1990 x Annex B x Years			0.031	0.031
			(0.019)	(0.019)
Annex B x Years			-0.014**	-0.014*
			(0.007)	(0.007)
Years			0.006***	0.006***
			(0.001)	(0.001)
Year FE	Yes	Yes	No	No
Molecule FE	Yes	Yes	Yes	Yes
Bootstraped	No	Yes	No	Yes
R-squared	0.989	0.989	0.987	0.987
Observations	864	864	864	864

Standard errors in parentheses
Dependent variable: Log count of patents
Years are relative to 1990.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Patents: Diff-in-Diff. Additional Specifications

	(1)	(2)	(3)	(4)
Post 1990 x Annex B	-0.251***	-0.251***		
	(0.065)	(0.059)		
Post 1990 x Annex B x Years			-0.040**	-0.040***
			(0.017)	(0.015)
Annex B x Years			-0.004	-0.004
			(0.005)	(0.006)
Years			0.026***	0.026***
			(0.002)	(0.002)
Year FE	Yes	Yes	No	No
Molecule FE	Yes	Yes	Yes	Yes
Bootstraped	No	Yes	No	Yes
R-squared	0.968	0.968	0.967	0.967
Observations	840	840	840	840

Standard errors in parentheses
Dependent variable: Log count of articles
Years are relative to 1990.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Articles: Diff-in-Diff. Additional Specifications

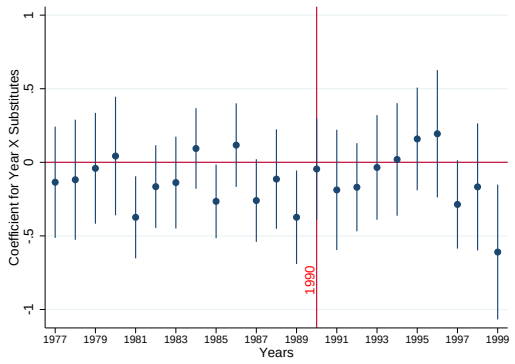


Figure 16: Patents: Differences-in-difference treatment effects by year

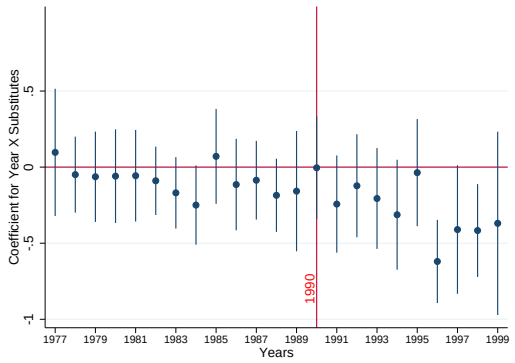


Figure 17: Articles: Differences-in-difference treatment effects by year

	(1)	(2)	(3)	(4)	(5)	(6)
Post 1990 x Annex B	-0.002	-0.091**	-0.093**			
	(0.061)	(0.041)	(0.042)			
Weighted mean proportion of topic 1		0.750**			0.428	
		(0.370)			(0.360)	
Weighted mean proportion of topic 2		1.379***			1.603***	
		(0.430)			(0.486)	
Weighted mean proportion of topic 3		0.708***			0.775***	
		(0.218)			(0.215)	
Weighted mean proportion of topic 4		0.937**			1.048***	
		(0.394)			(0.382)	
Weighted mean proportion of topic 5		1.481**			1.436**	
		(0.600)			(0.609)	
Mean proportion of topic 1			0.572			0.137
			(0.412)			(0.383)
Mean proportion of topic 2			1.509***			1.721***
			(0.455)			(0.520)
Mean proportion of topic 3			0.744***			0.841***
			(0.246)			(0.236)
Mean proportion of topic 4			0.779*			0.959**
			(0.420)			(0.391)
Mean proportion of topic 5			1.856**			1.814**
			(0.932)			(0.903)
Post 1990 x Annex B x Years				0.031	-0.007	-0.009
				(0.019)	(0.012)	(0.012)
Annex B x Years				-0.014**	-0.006	-0.005
				(0.007)	(0.004)	(0.004)
Years				0.006***	0.006***	0.006***
				(0.001)	(0.001)	(0.001)
Year FE	Yes	Yes	Yes	No	No	No
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.989	0.994	0.994	0.987	0.992	0.992
Observations	864	864	864	864	864	864

Standard errors in parentheses
Dependent variable: Log count of patents
Years are relative to 1990.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Patents: Diff-in-Diff. Controlling for topic proportions.

	(1)	(2)	(3)	(4)	(5)	(6)
Post 1990 x Annex B	-0.251***	-0.264***	-0.260***			
	(0.065)	(0.056)	(0.056)			
Weighted mean proportion of topic 1		1.660***			1.823***	
		(0.509)			(0.518)	
Weighted mean proportion of topic 2		0.342			0.391	
		(0.464)			(0.448)	
Weighted mean proportion of topic 3		1.328**			1.104*	
		(0.605)			(0.595)	
Weighted mean proportion of topic 4		0.545			0.454	
		(0.434)			(0.440)	
Weighted mean proportion of topic 5		0.818*			0.872**	
		(0.430)			(0.436)	
Mean proportion of topic 1			1.783***			2.063***
			(0.604)			(0.609)
Mean proportion of topic 2			0.266			0.191
			(0.602)			(0.586)
Mean proportion of topic 3			1.096			0.799
			(0.793)			(0.756)
Mean proportion of topic 4			0.715			0.684
			(0.586)			(0.595)
Mean proportion of topic 5			0.790			0.932*
			(0.541)			(0.551)
Post 1990 x Annex B x Years				-0.040**	-0.044***	-0.044***
				(0.017)	(0.015)	(0.015)
Annex B x Years				-0.004	-0.004	-0.003
				(0.005)	(0.005)	(0.005)
Years				0.026***	0.024***	0.024***
				(0.002)	(0.002)	(0.002)
Year FE	Yes	Yes	Yes	No	No	No
Molecule FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.968	0.974	0.974	0.967	0.973	0.973
Observations	840	840	840	840	840	840

Standard errors in parentheses
Dependent variable: Log count of articles
Years are relative to 1990.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Articles: Diff-in-Diff. Controlling for topic proportions.

4 Topic Proportions

- Table 18 shows the composition of the 5 topics in articles.
- Table 19 shows the composition of the 5 topics in patents.

Topic 1 - words	Topic 1 - prob	Topic 2 - words	Topic 2 - prob	Topic 3 - words	Topic 3 - prob
alderson	0.0110	xining	0.0074	sciatica	0.0064
paration	0.0062	sciatica	0.0058	leontein	0.0062
tano	0.0048	allot	0.0056	99tc	0.0052
i97o	0.0042	references1part	0.0045	rsquared	0.0043
100with	0.0037	kiho	0.0043	alderson	0.0040
shoji	0.0035	proteinase	0.0043	linares	0.0037
govemed	0.0034	ec	0.0038	cubooctahedral	0.0036
g1cnac	0.0034	3glucan	0.0034	twoyearold	0.0036
leontein	0.0032	boundary	0.0030	diisothiocyanat...	0.0033
coherent	0.0030	bzr	0.0029	coherent	0.0031
iwt	0.0026	proteincarbohyd...	0.0028	whcn	0.0029
ozonedepletion	0.0026	iwt	0.0026	siltclay	0.0028
salvatore	0.0025	ridines	0.0026	suhl	0.0027
suhl	0.0025	ircicopph32	0.0024	volumic	0.0024
chouroulinkov	0.0025	esc	0.0023	reoxida	0.0024
paren	0.0025	linares	0.0022	lindsay	0.0024
linares	0.0024	a3d	0.0022	op2	0.0023
flawlessly	0.0023	pmc	0.0022	85for	0.0023
sciatica	0.0022	thetase	0.0022	i97o	0.0023
cyclohexanes	0.0022	leontein	0.0022	ozonedepletion	0.0022

Topic 4 - words	Topic 4 - prob	Topic 5 - words	Topic 5 - prob
g1cnac	0.0068	guczi	0.0084
c5h9	0.0053	g1cnac	0.0073
kiho	0.0050	dioctyltin	0.0062
nndimethylamide	0.0048	12as	0.0055
experiments15	0.0048	100with	0.0047
22hydroxy	0.0047	lpy	0.0044
xining	0.0046	linares	0.0039
penetrance	0.0046	i97o	0.0034
guczi	0.0045	volumic	0.0034
cyclohexanes	0.0042	10e6	0.0034
aldopyranose	0.0034	macek	0.0033
paren	0.0033	silurian	0.0031
esc	0.0028	octagonal	0.0031
100with	0.0028	crocodile	0.0030
peonai	0.0027	paren	0.0030
noakowskiego	0.0026	leontein	0.0028
me0h	0.0026	methylhexane	0.0027
paration	0.0025	whcn	0.0026
whcn	0.0025	z4	0.0026
6570degc	0.0025	heptanoic	0.0024

Figure 18: List of the 5 topics generated by the LDA model on articles.

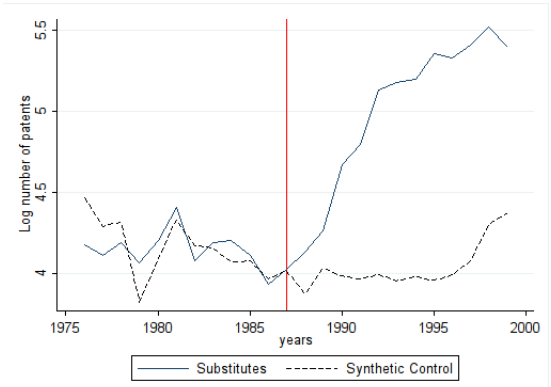
Topic 1 - words	Topic 1 - prob	Topic 2 - words	Topic 2 - prob	Topic 3 - words	Topic 3 - prob
crotononitrile	0.0090	andreu	0.0141	neal	0.0323
remote	0.0063	sulfon	0.0075	isopropyltrimet...	0.0276
dialkylhydantoi...	0.0047	phosphatidylino...	0.0072	inducers	0.0236
amineisocyanate	0.0043	maker	0.0060	amineprotecting	0.0180
petronate	0.0041	isopropyltrimet...	0.0058	heterophasic	0.0133
maker	0.0032	remote	0.0053	recrystallising	0.0119
sdspolyacrylami...	0.0031	satterfield	0.0049	flavipes	0.0090
sulfosuccinimid...	0.0030	recrystallising	0.0048	fmlp	0.0077
propanedioate	0.0030	neal	0.0042	tetraalkylphosp...	0.0071
highmoisture	0.0029	pertechnetate	0.0040	ocallaghan	0.0071
isocyanatomethy...	0.0028	brittleness	0.0040	topcoatings	0.0069
unseasoned	0.0028	crotononitrile	0.0037	photovolt	0.0045
hartog	0.0028	amineprotecting	0.0035	cocobetaine	0.0042
pathophysiologi...	0.0026	drifting	0.0034	ceratophyllus	0.0039
fluorimetric	0.0025	dihomogammalino...	0.0033	nitrofurantoin	0.0037
methylsulfonyle...	0.0024	inducers	0.0032	berkshire	0.0036
behaving	0.0024	dialkylhydantoi...	0.0030	wallenfels	0.0036
esterethers	0.0023	ciganek	0.0030	nnotriglycidyl	0.0035
varicose	0.0023	styreneglycidyl...	0.0030	genbank	0.0035
dissoluble	0.0022	highmoisture	0.0029	carboxyphenoxy	0.0034

Topic 4 - words	Topic 4 - prob	Topic 5 - words	Topic 5 - prob
topcoatings	0.0071	trisethyl	0.0157
heterophasic	0.0054	maker	0.0128
neal	0.0052	amineprotecting	0.0066
maker	0.0052	conservation	0.0064
inducers	0.0048	neal	0.0063
amineprotecting	0.0045	inducers	0.0059
recrystallising	0.0043	satterfield	0.0055
ceratophyllus	0.0042	recrystallising	0.0055
tripod	0.0041	alkoxyiminoc	0.0050
costabilisers	0.0037	delgado	0.0048
cconr	0.0037	isopropyltrimet...	0.0046
varicose	0.0034	heterophasic	0.0045
isopropyltrimet...	0.0032	brittleness	0.0043
satterfield	0.0032	ceftazidime	0.0041
aminco	0.0030	ocallaghan	0.0041
polycations	0.0028	ceratophyllus	0.0037
carnuba	0.0027	photovolt	0.0037
mutation	0.0027	hydroxyquinone	0.0037
holder	0.0026	polycations	0.0032
brittleness	0.0026	gequivalent	0.0031

Figure 19: List of the 5 topics generated by the LDA model on patents.

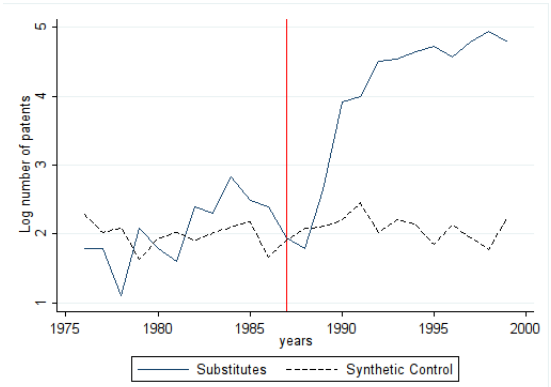
5 SCM

- Figure 21 displays the results of the synthetic control method for OSD-substitutes with 5 LDA topics for the strong, intermediate and weak criteria for assigning articles to treatment groups.
- Figure 20 displays the results of the synthetic control method for OSD-substitutes with 5 LDA topics for the strong, intermediate and weak criteria for assigning articles to treatment groups.
- Figure 23 displays the results of the synthetic control method for Annex A compounds with 5 LDA topics for the strong, intermediate and weak criteria for assigning articles to treatment groups.
- Figure 22 displays the results of the synthetic control method for Annex A compounds with 5 LDA topics for the strong, intermediate and weak criteria or assigning articles to treatment groups.
- Figure 25 displays the results of the synthetic control method for Annex B compounds with 5 LDA topics for the strong, intermediate and weak criteria for assigning articles to treatment groups.
- Figure 24 displays the results of the synthetic control method for Annex B compounds with 5 LDA topics for the strong, intermediate and weak criteria for assigning articles to treatment groups.



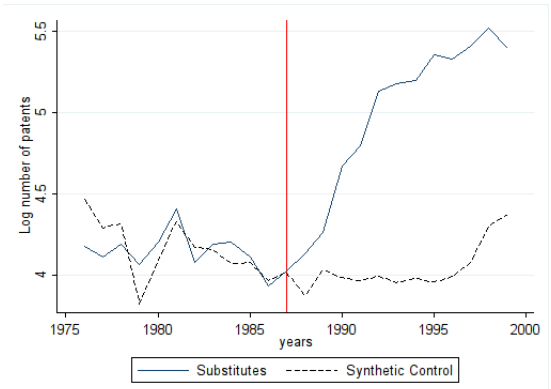
(a) Weak - Raw effect.

(b) Weak - Placebo tests.



(c) Intermediate - Raw effect.

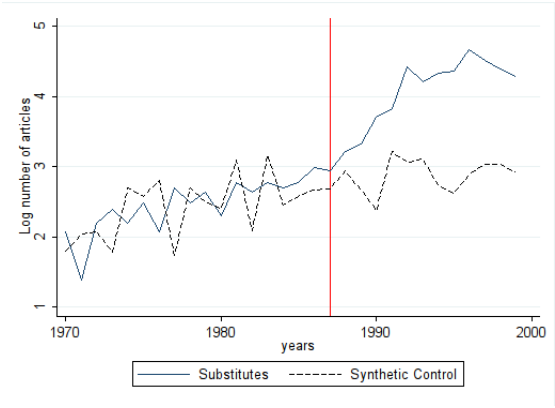
(d) Intermediate - Placebo tests.



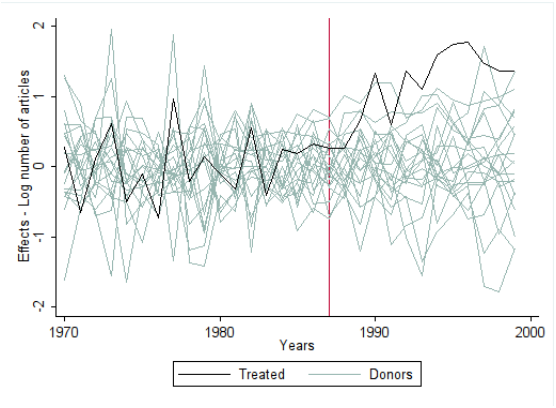
(e) Strong - Raw effect.

(f) Strong - Placebo tests.

Figure 20: Synthetic Control for Substitutes in Patents (LDA 5 topics, weighted means, small pool)



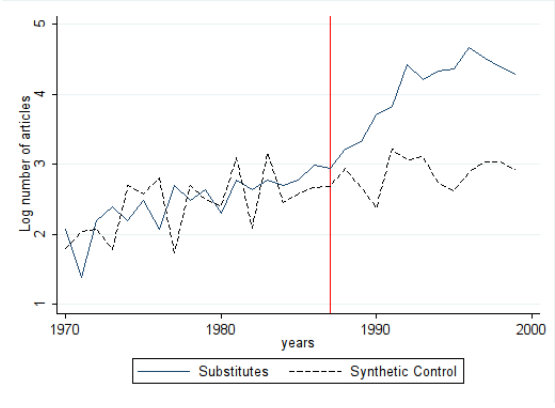
(a) Weak - Raw effect.



(b) Weak - Placebo tests.

(c) Intermediate - Raw effect.

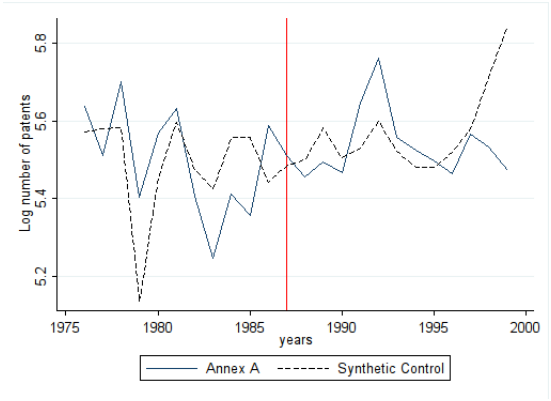
(d) Intermediate - Placebo tests.



(e) Strong - Raw effect.

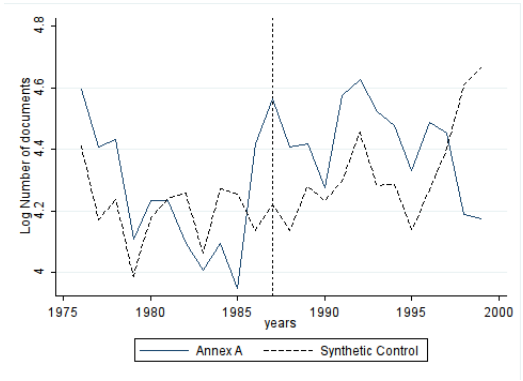
(f) Strong - Placebo tests.

Figure 21: Synthetic Control for Substitutes in Articles (LDA 5 topics, weighted means, small pool)



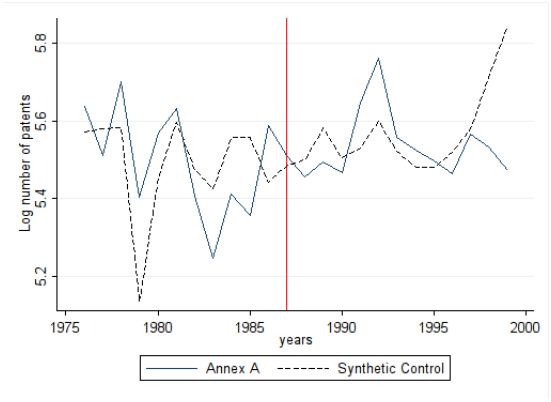
(a) Weak - Raw effect.

(b) Weak - Placebo tests.

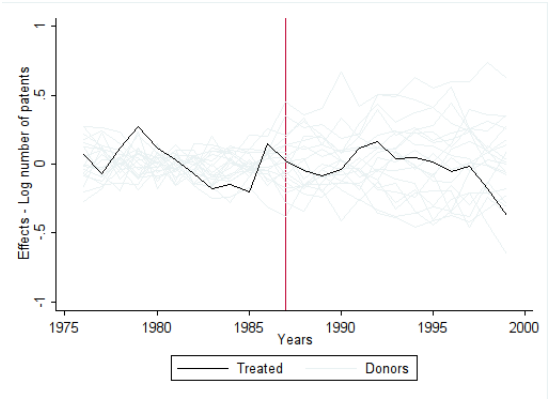


(c) Intermediate - Raw effect.

(d) Intermediate - Placebo tests.

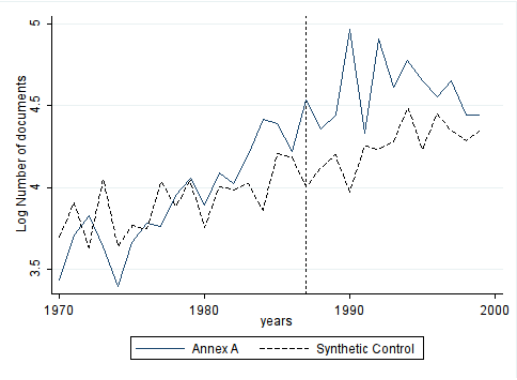


(e) Strong - Raw effect.



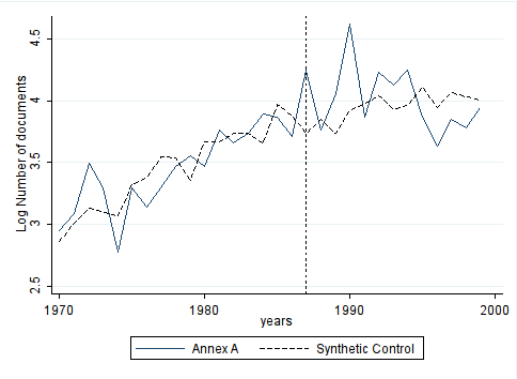
(f) Strong - Placebo tests.

Figure 22: Synthetic Control for Annex A Compounds in Patents (LDA 5 topics, weighted means, small pool)



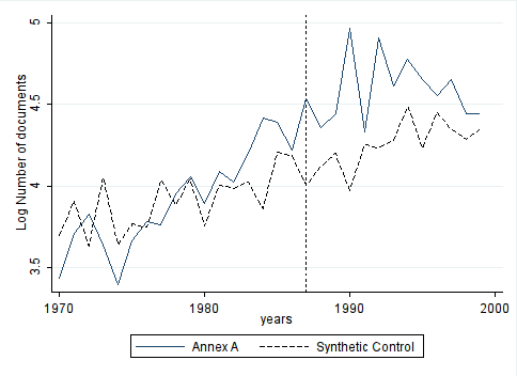
(a) Weak - Raw effect.

(b) Weak - Placebo tests.



(c) Intermediate - Raw effect.

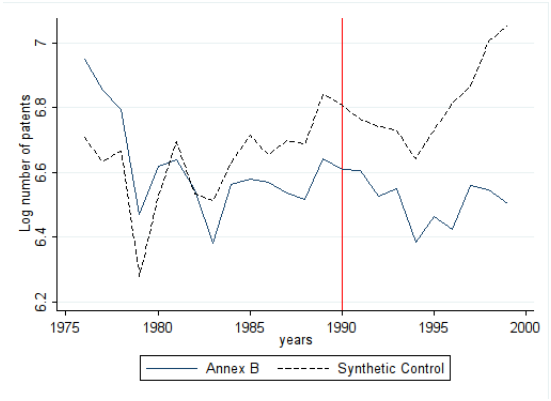
(d) Intermediate - Placebo tests.



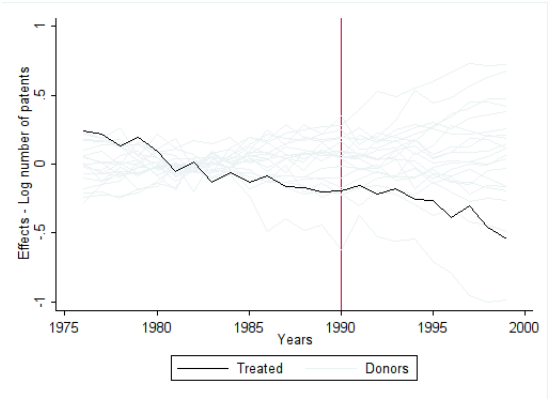
(e) Strong - Raw effect.

(f) Strong - Placebo tests.

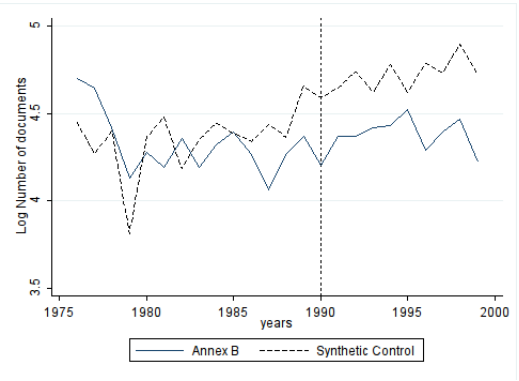
Figure 23: Synthetic Control for Annex A Compounds in Articles (LDA 5 topics, weighted means, small pool)



(a) Weak - Raw effect.

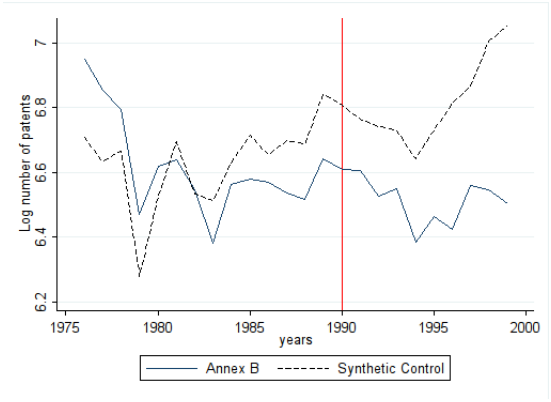


(b) Weak - Placebo tests.

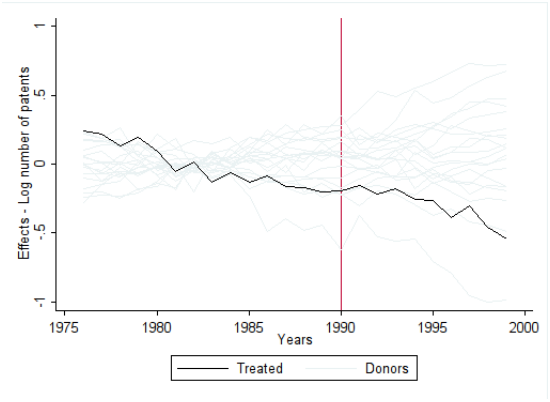


(c) Intermediate - Raw effect.

(d) Intermediate - Placebo tests.

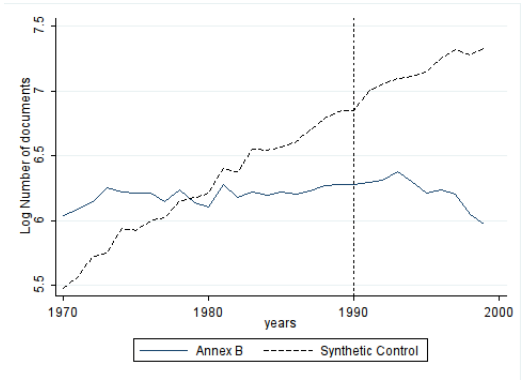


(e) Strong - Raw effect.

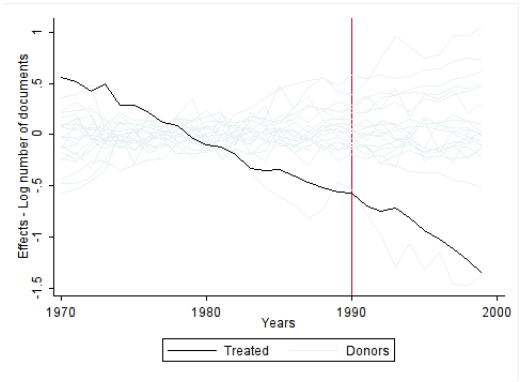


(f) Strong - Placebo tests.

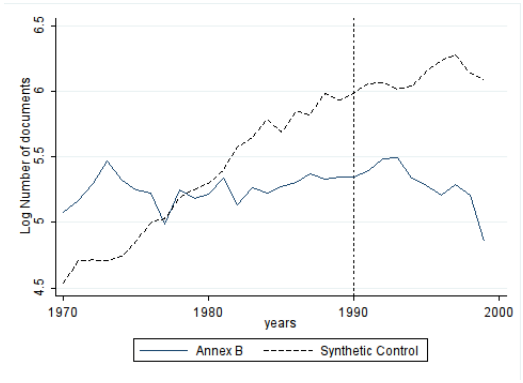
Figure 24: Synthetic Control for Annex B Compounds in Patents (LDA 5 topics, weighted means, small pool)



(a) Weak - Raw effect.

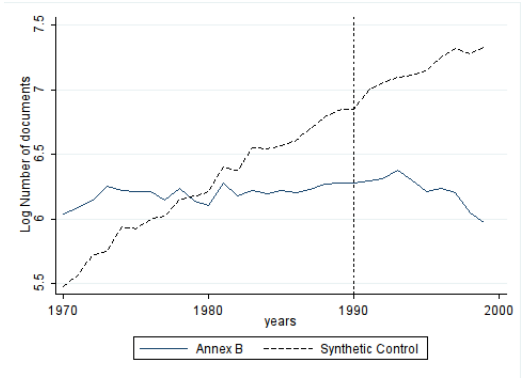


(b) Weak - Placebo tests.

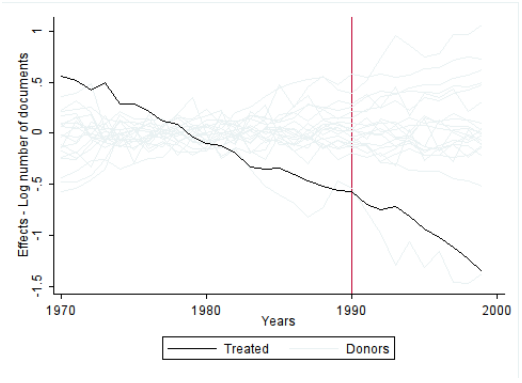


(c) Intermediate - Raw effect.

(d) Intermediate - Placebo tests.



(e) Strong - Raw effect.



(f) Strong - Placebo tests.

Figure 25: Synthetic Control for Annex B Compounds in Articles (LDA 5 topics, weighted means, small pool)

6 Robustness Checks

- Figure 26 plots the number of documents per year per group and per method.
- Table 18 displays the number of articles per group and assignment method.
- Table 19 displays the number of patents per group and assignment method.
- Table 20 displays the performance Summary of Several SCM Implementations for Patents Assuming Anticipation.
- Table 21 displays the performance Summary of Several SCM Implementations for Articles Assuming Anticipation.
- Table 22 displays the performance Summary of Several SCM Implementations for Patents, with counts (not log).
- Table 23 displays the performance Summary of Several SCM Implementations for Articles, with count (not log).
- Table 24 displays the performance Summary of Several SCM Implementations for Patents, with count (not log), for Annex A.
- Table 25 displays the performance Summary of Several SCM Implementations for Articles, with count (not log), for Annex A.
- Table 26 displays the performance Summary of Several SCM Implementations for Patents, with count (not log), for Annex B.
- Table 27 displays the performance Summary of Several SCM Implementations for Articles, with count (not log), for Annex B.
- Table 28 displays the performance Summary of Several SCM Implementations for Patents, without log.
- Table 29 displays the performance summary of Several SCM Implementations for Articles, without log.
- Table 30 displays the performance Summary of Several SCM Implementations for Patents, without log, for Annex A.
- Table 31 displays the performance Summary of Several SCM Implementations for Articles, without log, for Annex A.
- Table 32 displays the performance Summary of Several SCM Implementations for Patents, without log, for Annex B.

- Table 33 displays the performance Summary of Several SCM Implementations for Articles, without log, for Annex B.

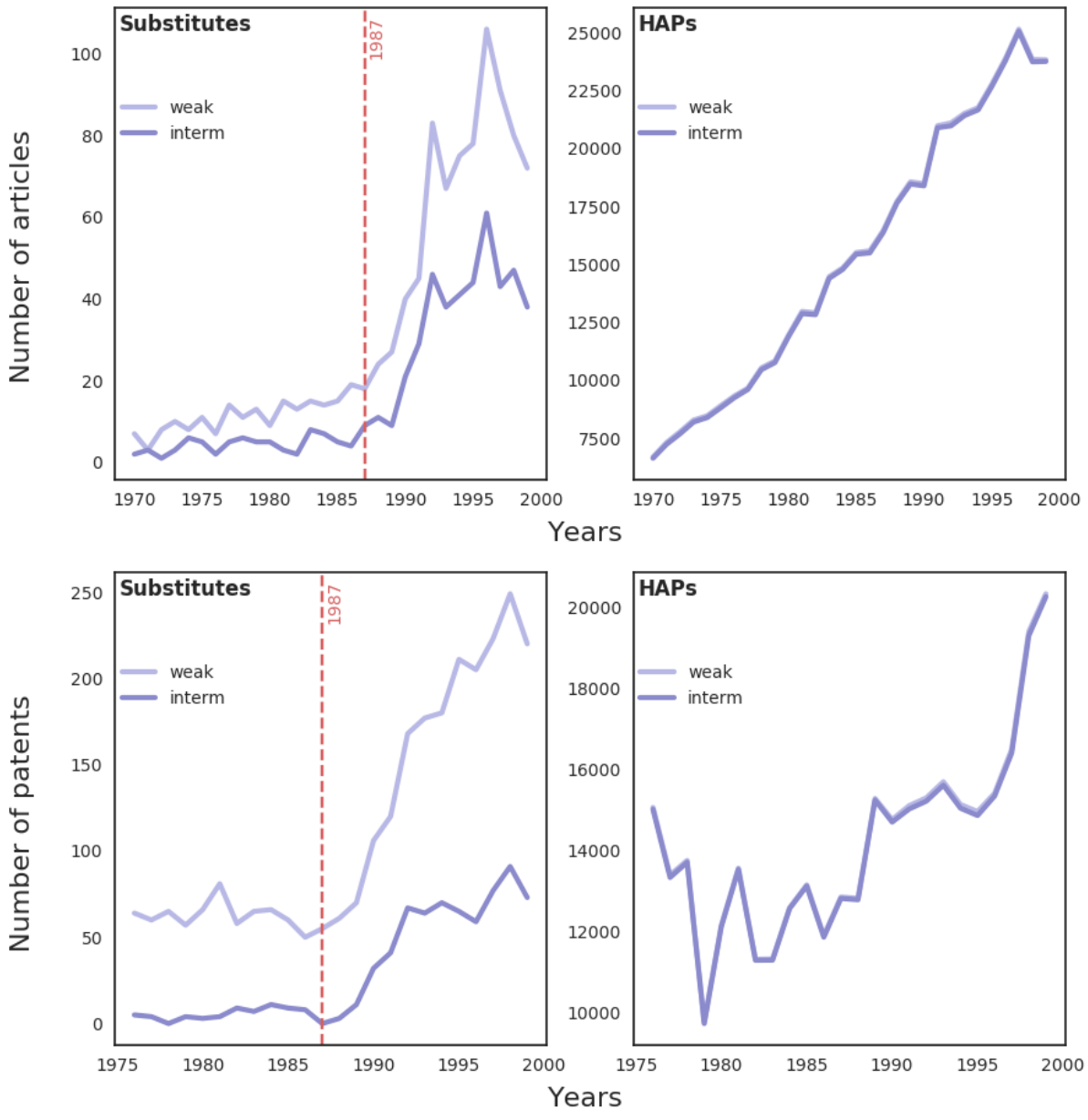


Figure 26: Counts per year per group and per method in articles and patents.

	A strong	A interm	A weak	B strong	B interm	B weak	S strong	S interm	S weak	H strong	H interm	H weak
1970	10	15	30	52	102	417	2	2	7	6313	6637	6692
1971	11	18	40	57	127	438	2	3	3	6912	7238	7315
1972	18	31	45	64	143	465	1	1	8	7359	7693	7784
1973	18	24	37	76	158	517	3	3	10	7833	8200	8287
1974	10	14	29	65	138	501	5	6	8	8023	8393	8471
1975	12	24	38	66	145	499	4	5	11	8455	8815	8904
1976	10	21	43	60	130	499	2	2	7	8866	9248	9325
1977	14	22	42	52	102	464	3	5	14	9234	9605	9663
1978	21	29	51	61	132	510	5	6	11	10070	10459	10538
1979	23	28	57	60	127	460	3	5	13	10406	10764	10837
1980	21	28	48	65	125	446	2	5	9	11547	11882	11950
1981	26	39	59	73	158	533	2	3	15	12485	12880	12975
1982	27	36	55	59	134	481	2	2	13	12466	12833	12915
1983	23	38	66	80	143	502	5	8	15	14025	14397	14473
1984	32	43	82	73	128	488	5	7	14	14397	14785	14850
1985	30	45	80	84	146	503	4	5	15	15063	15442	15516
1986	23	36	67	74	159	493	4	4	19	15136	15497	15589
1987	40	64	92	104	169	506	7	9	18	16032	16384	16472
1988	28	39	77	89	147	529	5	11	24	17231	17644	17715
1989	29	52	84	93	167	532	5	9	27	18076	18470	18564
1990	67	92	143	93	164	532	7	21	40	17998	18394	18489
1991	34	45	75	100	173	539	20	29	45	20505	20898	20986
1992	40	64	134	86	183	547	22	46	83	20564	20983	21110
1993	38	53	100	111	184	588	24	38	67	20984	21428	21522
1994	40	67	118	99	164	543	21	41	75	21248	21670	21767
1995	26	44	104	67	138	496	22	44	78	22263	22672	22773
1996	14	31	94	68	134	508	30	61	106	23342	23776	23875
1997	25	40	104	64	135	494	29	43	91	24622	25062	25154
1998	24	39	84	54	128	420	25	47	80	23403	23746	23851
1999	21	45	84	47	92	392	24	38	72	23421	23761	23833

Table 18: Number of articles per molecule groups.

	A interm	A weak	B interm	B weak	S interm	S weak	H interm	H weak
1976	59	280	33	1040	5	64	15009	15057
1977	56	246	29	945	4	60	13331	13393
1978	51	298	27	892	0	65	13704	13751
1979	40	221	19	645	4	57	9733	9771
1980	39	261	24	747	3	66	12120	12153
1981	38	278	24	764	4	81	13527	13563
1982	42	222	29	695	9	58	11291	11318
1983	38	189	18	589	7	65	11294	11325
1984	29	223	17	708	11	66	12567	12588
1985	35	211	20	720	9	60	13120	13149
1986	53	266	20	712	8	50	11860	11895
1987	65	246	22	689	0	55	12814	12850
1988	56	233	22	676	3	61	12782	12822
1989	53	242	28	766	11	70	15238	15276
1990	51	236	23	740	32	106	14695	14758
1991	71	282	30	738	41	120	15018	15101
1992	64	317	26	682	67	168	15210	15292
1993	65	258	33	699	64	177	15603	15689
1994	55	250	35	591	70	180	15037	15128
1995	48	243	50	641	65	211	14856	14954
1996	49	235	33	616	59	205	15337	15406
1997	64	260	24	705	77	223	16398	16492
1998	42	252	38	695	91	249	19285	19387
1999	44	237	39	667	73	220	20244	20318

Table 19: Number of patents per molecule groups.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
weak	unweighted	whole sample	0.10	0.01	0.64	1991
weak	weighted	small pool	0.17	0.00	1.03	1991
weak	weighted	whole sample	0.18	0.04	0.83	1990
weak	unweighted	small pool	0.22	0.00	1.07	1991

Table 20: Performance Summary of Several SCM Implementations for Patents Assuming Anticipation.

Notes: For these experiments, the treatment year is redefined as 1985 and the synthetic control constructed using data until 1983. The table is ranked according to the magnitude of the pre-RMPSE with smaller pre-RMPSE at the top of the table. The lower the pre-RMPSE, the better the fit between the synthetic control and the treated unit over pre-treatment years. Hence, the lower the pre-RMPSE, the more credible it is that the synthetic control appropriately proxies the counterfactual. The best implementation seems to call for using Log Count from the intermediate rule as outcome variable, weighted topic proportions as covariates, and the whole sample of HAPs as donor pool. "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count. Since the RMSPE calculated on a logarithmized variable is necessarily smaller than when the variable is not logarithmized, I compute the log value of the Pre RMSPEs when the outcomes are the non-logged counts so to provide a way to gauge magnitude differences with the logged counts.

Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
unweighted	whole sample	0.34	0.029	1.25	1990
weighted	whole sample	0.36	0.119	1.01	1990
unweighted	small pool	0.45	0.050	1.07	1990
weighted	small pool	0.48	0.050	1.25	1990

Table 21: Performance Summary of Several SCM Implementations for Articles Assuming Anticipation.
Notes: For these experiments, the treatment year is redefined as 1985 and the synthetic control constructed using data until 1983. The table is ranked according to the magnitude of the pre-RMPSE with smaller pre-RMPSE at the top of the table. The lower the pre-RMPSE, the better the fit between the synthetic control and the treated unit over pre-treatment years. Hence, the lower the pre-RMPSE, the more credible it is that the synthetic control appropriately proxies the counterfactual. The best implementation seems to call for using Log Count from the intermediate rule as outcome variable, weighted topic proportions as covariates, and the whole sample of HAPs as donor pool. "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count. Since the RMSPE calculated on a logarithmized variable is necessarily smaller than when the variable is not logarithmized, I compute the log value of the Pre RMSPEs when the outcomes are the non-logged counts so to provide a way to gauge magnitude differences with the logged counts.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
intermediate	unweighted	small pool	3.54	0.00	72.05	1990
intermediate	weighted	small pool	3.57	0.00	71.77	1990
intermediate	weighted	whole sample	3.85	0.00	71.24	1990
intermediate	unweighted	whole sample	4.25	0.01	71.88	1990
weak	unweighted	whole sample	8.12	0.00	84.52	1992
weak	weighted	whole sample	8.41	0.00	100.92	1992
weak	weighted	small pool	8.52	0.00	110.45	1992
weak	unweighted	small pool	12.49	0.00	113.09	1992

Table 22: Performance Summary of Several SCM Implementations for Patents, with counts (not log).
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
intermediate	weighted	small pool	1.54	0.000	35.17	1990
intermediate	unweighted	small pool	1.78	0.000	35.52	1990
intermediate	unweighted	whole sample	1.96	0.000	35.04	1992
intermediate	weighted	whole sample	2.03	0.000	35.39	1992
weak	weighted	small pool	2.65	0.000	44.08	1990
weak	unweighted	small pool	2.86	0.000	42.6	1990
weak	weighted	whole sample	4.44	0.006	45.81	1998
weak	unweighted	whole sample	5.65	0.006	46.73	1994

Table 23: Performance Summary of Several SCM Implementations for Articles, with count (not log).
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
intermediate	weighted	small pool	12.50	0.55	-4.06	–
intermediate	unweighted	whole sample	13.24	0.43	-10.01	–
intermediate	weighted	whole sample	13.70	0.42	-9.67	–
intermediate	unweighted	small pool	14.12	0.70	-1.46	–
weak	unweighted	whole sample	22.96	0.55	-15.9	–
weak	weighted	whole sample	23.17	0.55	-9.99	–
weak	weighted	small pool	31.34	0.80	-12.52	–
weak	unweighted	small pool	32.33	0.60	-27.3	–

Table 24: Performance Summary of Several SCM Implementations for Patents, with count (not log), for Annex A.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
intermediate	unweighted	whole sample	6.93	0.218	-9.31	1990
intermediate	weighted	whole sample	7.36	0.200	-10.02	1990
weak	weighted	whole sample	8.02	0.260	4.25	1990
intermediate	unweighted	small pool	8.24	0.300	-4.43	1990
weak	unweighted	whole sample	8.27	0.276	2.15	1990
weak	unweighted	small pool	8.63	0.250	-10.03	–
intermediate	weighted	small pool	8.72	0.250	2.71	1990
weak	weighted	small pool	15.51	0.700	20.58	1990

Table 25: Performance Summary of Several SCM Implementations for Articles, with count (not log), for Annex A.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
intermediate	unweighted	small pool	13.21	0.45	-19.0	–
intermediate	unweighted	whole sample	13.50	0.45	-17.5	–
intermediate	weighted	whole sample	14.37	0.38	-19.42	–
intermediate	weighted	small pool	18.77	0.45	-30.44	–
weak	weighted	whole sample	65.39	0.29	-131.54	–
weak	unweighted	whole sample	66.32	0.29	-123.98	–
weak	unweighted	small pool	98.83	0.25	-199.65	–
weak	weighted	small pool	122.56	0.35	-243.17	–

Table 26: Performance Summary of Several SCM Implementations for Patents, with count (not log), for Annex B.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
intermediate	weighted	whole sample	37.09	0.161	-87.39	–
intermediate	unweighted	whole sample	41.97	0.166	-101.2	–
intermediate	unweighted	small pool	78.55	0.550	-197.83	–
weak	unweighted	whole sample	84.71	0.129	-248.24	–
weak	weighted	whole sample	90.61	0.169	-243.75	–
intermediate	weighted	small pool	97.29	0.600	-247.15	–
weak	weighted	small pool	191.98	0.300	-655.45	–
weak	unweighted	small pool	197.01	0.400	-640.63	–

Table 27: Performance Summary of Several SCM Implementations for Articles, with count (not log), for Annex B.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Outcome	Pre RMSPE	p-value	ATE	Year
weak	weighted	DD subset	Log Count	0.12	0.00	0.92	1990
weak	unweighted	DD subset	Log Count	0.15	0.00	0.87	1990
intermediate	weighted	DD subset	Log Count	0.45	0.10	1.67	1989
intermediate	unweighted	DD subset	Log Count	0.46	0.10	1.68	1989
intermediate	weighted	DD subset	Count	3.70	0.02	67.43	1990
intermediate	unweighted	DD subset	Count	3.86	0.02	67.48	1990
weak	weighted	DD subset	Count	6.42	0.00	101.53	1992
weak	unweighted	DD subset	Count	8.06	0.00	82.73	1992

Table 28: Performance Summary of Several SCM Implementations for Patents, without log.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Outcome	Pre RMSPE	p-value	ATE	Year
intermediate	unweighted	DD subset	Log Count	0.40	0.025	1.03	1992
intermediate	weighted	DD subset	Log Count	0.44	0.025	0.99	1990
weak	weighted	DD subset	Log Count	0.48	0.262	0.79	1990
weak	unweighted	DD subset	Log Count	0.51	0.366	0.72	1990
intermediate	unweighted	DD subset	Count	2.82	0.000	30.63	1992
intermediate	weighted	DD subset	Count	3.97	0.000	27.59	1992
weak	unweighted	DD subset	Count	5.37	0.000	31.68	–
weak	weighted	DD subset	Count	5.41	0.000	33.68	–

Table 29: Performance Summary of Several SCM Implementations for Articles, without log.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Outcome	Pre RMSPE	p-value	ATE	Year
weak	unweighted	DD subset	Log Count	0.16	0.75	0.2	1992
weak	weighted	DD subset	Log Count	0.16	1.00	-0.09	–
intermediate	weighted	DD subset	Log Count	0.23	0.67	0.01	–
intermediate	unweighted	DD subset	Log Count	0.23	0.75	-0.07	–
intermediate	weighted	DD subset	Count	14.55	0.33	-28.8	–
intermediate	unweighted	DD subset	Count	15.86	0.88	-3.67	–
weak	weighted	DD subset	Count	31.62	0.62	-31.27	–
weak	unweighted	DD subset	Count	32.47	0.75	-21.03	–

Table 30: Performance Summary of Several SCM Implementations for Patents, without log, for Annex A.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Outcome	Pre RMSPE	p-value	ATE	Year
intermediate	weighted	DD subset	Log Count	0.31	0.333	0.37	1990
weak	unweighted	DD subset	Log Count	0.31	0.500	0.24	1990
weak	weighted	DD subset	Log Count	0.32	0.250	0.64	1990
intermediate	unweighted	DD subset	Log Count	0.32	0.348	0.23	1990
intermediate	weighted	DD subset	Count	9.40	0.375	11.55	1990
intermediate	unweighted	DD subset	Count	9.96	0.348	11.49	1990
weak	unweighted	DD subset	Count	13.21	0.458	11.76	–
weak	weighted	DD subset	Count	13.59	0.333	26.28	1990

Table 31: Performance Summary of Several SCM Implementations for Articles, without log, for Annex A.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Outcome	Pre RMSPE	p-value	ATE	Year
weak	unweighted	DD subset	Log Count	0.12	0.44	-0.21	–
weak	weighted	DD subset	Log Count	0.13	0.30	-0.24	–
intermediate	unweighted	DD subset	Log Count	0.17	0.39	-0.22	–
intermediate	weighted	DD subset	Log Count	0.24	0.82	-0.2	–
intermediate	unweighted	DD subset	Count	14.03	0.50	-20.97	–
intermediate	weighted	DD subset	Count	17.08	0.50	-18.95	–
weak	weighted	DD subset	Count	114.58	0.33	-216.19	–
weak	unweighted	DD subset	Count	116.97	0.37	-205.94	–

Table 32: Performance Summary of Several SCM Implementations for Patents, without log, for Annex B.
Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

Rule	Topic Means	Donor Pool	Outcome	Pre RMSPE	p-value	ATE	Year
intermediate	weighted	DD subset	Log Count	0.22	0.000	-0.54	–
intermediate	unweighted	DD subset	Log Count	0.22	0.067	-0.54	–
weak	unweighted	DD subset	Log Count	0.28	0.667	-0.08	–
weak	weighted	DD subset	Log Count	0.28	0.857	-0.08	–
intermediate	unweighted	DD subset	Count	40.47	0.200	-120.52	–
intermediate	weighted	DD subset	Count	41.99	0.267	-133.11	–
weak	weighted	DD subset	Count	116.41	0.905	-44.25	1988
weak	unweighted	DD subset	Count	116.41	0.905	-44.25	1988

Table 33: Performance Summary of Several SCM Implementations for Articles, without log, for Annex B.

Notes: "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count.

7 Annex A and B compounds

- Table **??** displays the performance summary of the main SCM implementations for patents for Annex A.
- Table 35 displays the performance summary of the main SCM implementations for articles for Annex A.
- Table 36 displays the performance summary of the main SCM implementations for patents for Annex B.
- Table 37 displays the performance summary of the main SCM implementations for articles for Annex B.
- Figure 27 displays the SCM graphs for Annex A compounds using the weak rule (LDA 5 topics)
- Figure 28 displays the SCM graphs for Annex B compounds using the weak rule (LDA 5 topics, unweighted means)

Table 34: Performance summary of the main SCM implementations for patents for Annex A.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
weak	weighted	whole sample	0.10	0.43	-0.0	–
weak	unweighted	whole sample	0.11	0.51	-0.01	–
weak	unweighted	small pool	0.14	0.80	-0.09	–
weak	weighted	small pool	0.14	0.85	-0.03	–
intermediate	unweighted	whole sample	0.19	0.60	-0.02	–
intermediate	weighted	whole sample	0.19	0.62	-0.01	–
intermediate	weighted	small pool	0.20	0.55	0.07	–
intermediate	unweighted	small pool	0.21	0.70	0.09	–

Notes: The table is ranked according to the magnitude of the pre-RMPSE with smaller pre-RMPSE at the top of the table. The lower the pre-RMPSE, the better the fit between the synthetic control and the treated unit over pre-treatment years. Hence, the lower the pre-RMPSE, the more credible it is that the synthetic control appropriately proxies the counterfactual. The best implementation seems to call for using Log Count from the weak rule as outcome variable. Using weighted or unweigthed topic proportions as covariates does change much the pre-RMSPE. Using the whole sample of HAPs as donor pool or only the diff-in-diff subset yields similar results. The p-values are the best fitted synthetic controls indicate significance at the 95% level since p-values are smaller than 0.05. "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count. Since the RMSPE calculated on a logarithmized variable is necessarily smaller than when the variable is not logarithmized, I compute the log value of the Pre RMSPEs when the outcomes are the non-logged counts so to provide a way to gauge magnitude differences with the logged counts.

Table 35: Performance summary of the main SCM implementations for articles for Annex A.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
weak	unweighted	small pool	0.15	0.250	-0.12	–
weak	unweighted	whole sample	0.15	0.471	0.08	–
weak	weighted	whole sample	0.16	0.296	0.18	–
intermediate	unweighted	small pool	0.21	0.550	-0.03	1990
intermediate	weighted	small pool	0.22	0.200	0.04	1990
intermediate	unweighted	whole sample	0.22	0.395	0.01	1990
intermediate	weighted	whole sample	0.24	0.435	0.05	1990
weak	weighted	small pool	0.25	0.400	0.33	1990

Notes: The table is ranked according to the magnitude of the pre-RMPSE with smaller pre-RMPSE at the top of the table. The lower the pre-RMPSE, the better the fit between the synthetic control and the treated unit over pre-treatment years. Hence, the lower the pre-RMPSE, the more credible it is that the synthetic control appropriately proxies the counterfactual. The best implementation seems to call for using Log Count from the intermediate rule as outcome variable, weighted topic proportions as covariates, and the whole sample of HAPs as donor pool. "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count. Since the RMSPE calculated on a logarithmized variable is necessarily smaller than when the variable is not logarithmized, I compute the log value of the Pre RMSPEs when the outcomes are the non-logged counts so to provide a way to gauge magnitude differences with the logged counts.

Table 36: Performance summary of the main SCM implementations for patents for Annex B.

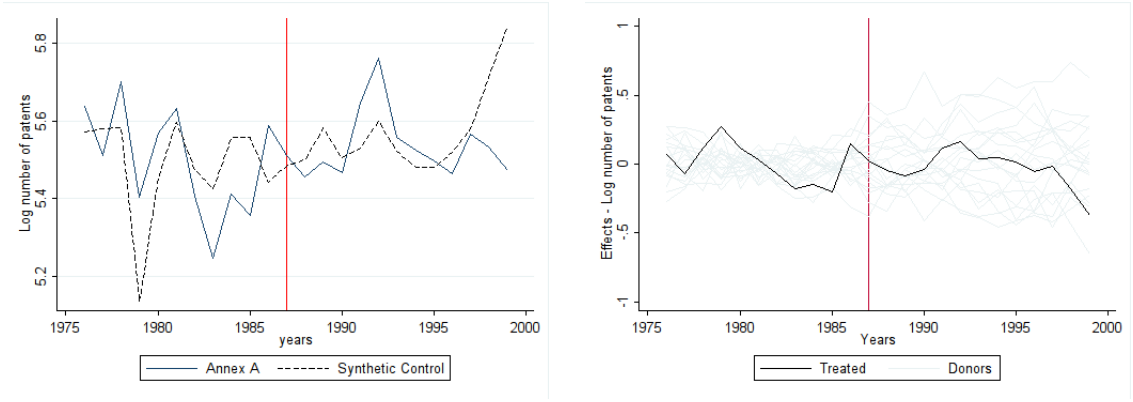
Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
weak	weighted	whole sample	0.08	0.14	-0.18	–
weak	unweighted	whole sample	0.08	0.17	-0.16	–
weak	unweighted	small pool	0.12	0.25	-0.23	–
weak	weighted	small pool	0.15	0.35	-0.28	–
intermediate	unweighted	small pool	0.16	0.45	-0.17	–
intermediate	unweighted	whole sample	0.16	0.58	-0.13	–
intermediate	weighted	whole sample	0.17	0.61	-0.14	–
intermediate	weighted	small pool	0.24	0.40	-0.32	–

Notes: The table is ranked according to the magnitude of the pre-RMPSE with smaller pre-RMPSE at the top of the table. The lower the pre-RMPSE, the better the fit between the synthetic control and the treated unit over pre-treatment years. Hence, the lower the pre-RMPSE, the more credible it is that the synthetic control appropriately proxies the counterfactual. The best implementation seems to call for using Log Count from the weak rule as outcome variable. Using weighted or unweigthed topic proportions as covariates does change much the pre-RMSPE. Using the whole sample of HAPs as donor pool or only the diff-in-diff subset yields similar results. The p-values are the best fitted synthetic controls indicate significance at the 95% level since p-values are smaller than 0.05. "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count. Since the RMSPE calculated on a logarithmized variable is necessarily smaller than when the variable is not logarithmized, I compute the log value of the Pre RMSPEs when the outcomes are the non-logged counts so to provide a way to gauge magnitude differences with the logged counts.

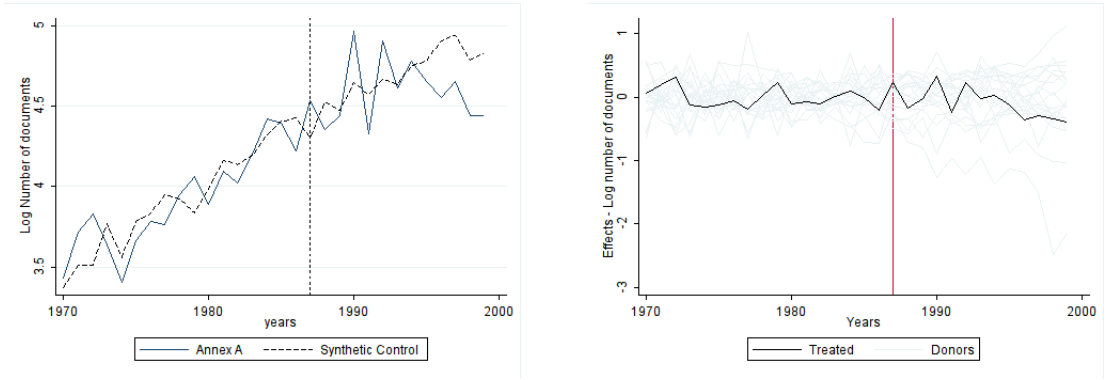
Table 37: Performance summary of the main SCM implementations for articles for Annex B.

Rule	Topic Means	Donor Pool	Pre RMSPE	p-value	ATE	Year
weak	unweighted	whole sample	0.15	0.036	-0.32	–
weak	weighted	whole sample	0.16	0.112	-0.28	–
intermediate	weighted	whole sample	0.18	0.188	-0.23	–
intermediate	unweighted	whole sample	0.19	0.077	-0.36	–
weak	weighted	small pool	0.38	0.200	-0.86	–
weak	unweighted	small pool	0.40	0.300	-0.86	–
intermediate	unweighted	small pool	0.43	0.250	-0.72	–
intermediate	weighted	small pool	0.46	0.350	-0.78	–

Notes: The table is ranked according to the magnitude of the pre-RMPSE with smaller pre-RMPSE at the top of the table. The lower the pre-RMPSE, the better the fit between the synthetic control and the treated unit over pre-treatment years. Hence, the lower the pre-RMPSE, the more credible it is that the synthetic control appropriately proxies the counterfactual. The best implementation seems to call for using Log Count from the intermediate rule as outcome variable, weighted topic proportions as covariates, and the whole sample of HAPs as donor pool. "Rule" corresponds to the rule used for assigning documents to molecule. "Donor Pool" indicates what sample of HAPs is used in the SCM procedure. For "subset", the sample of HAPs used corresponds to the subset found to have the pre-treatment slope closest to the treated unit. "Topic Means" indicates the procedure for aggregating the topic proportions at the molecule level. If "weighted", the calculated proportion of topic j for molecule i is the mean proportion of topic j across all documents mentioning molecule i, weighted by the number of times the molecule appears in the document. "Outcome" is either the count of documents, or the log count. Since the RMSPE calculated on a logarithmized variable is necessarily smaller than when the variable is not logarithmized, I compute the log value of the Pre RMSPEs when the outcomes are the non-logged counts so to provide a way to gauge magnitude differences with the logged counts.

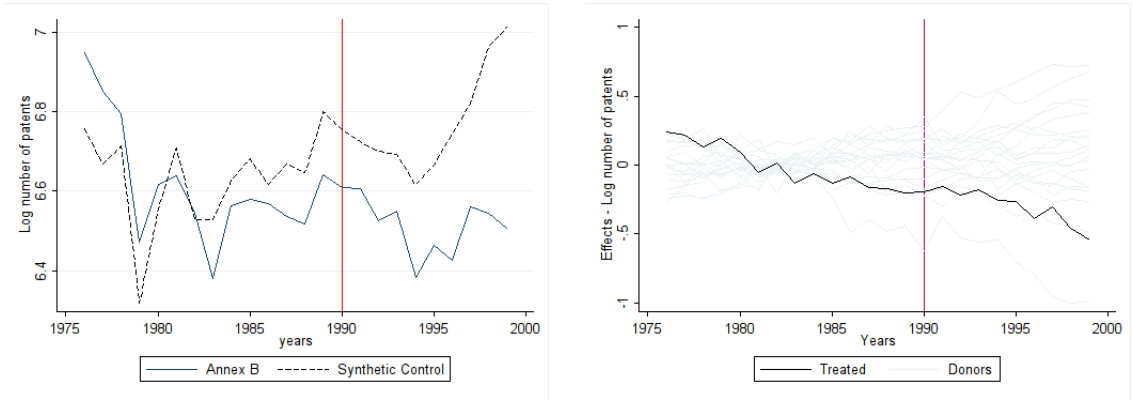


(a) Patents: raw effect (left) and placebo tests (right)



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

Figure 27: Synthetic control for Annex A compounds using the weak rule (LDA 5 topics)
Notes: Figures 27a and 27b display the results of the synthetic control method for Annex A compounds for patents and articles. In all cases, the method is implemented using the topic proportions of a LDA model with 5 topics and the weak rule for assigning documents to molecule groups. Weighted means of topic proportions are used for patent and unweighted means for articles because these are the specifications that yielded lowest pre-RMPSE. 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, the non-parametric tests to evaluate the significance of the results; black lines show the effect on the treated group relative to the control group, while each gray line is a placebo test performed on an unit drawn from the donor pool.



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

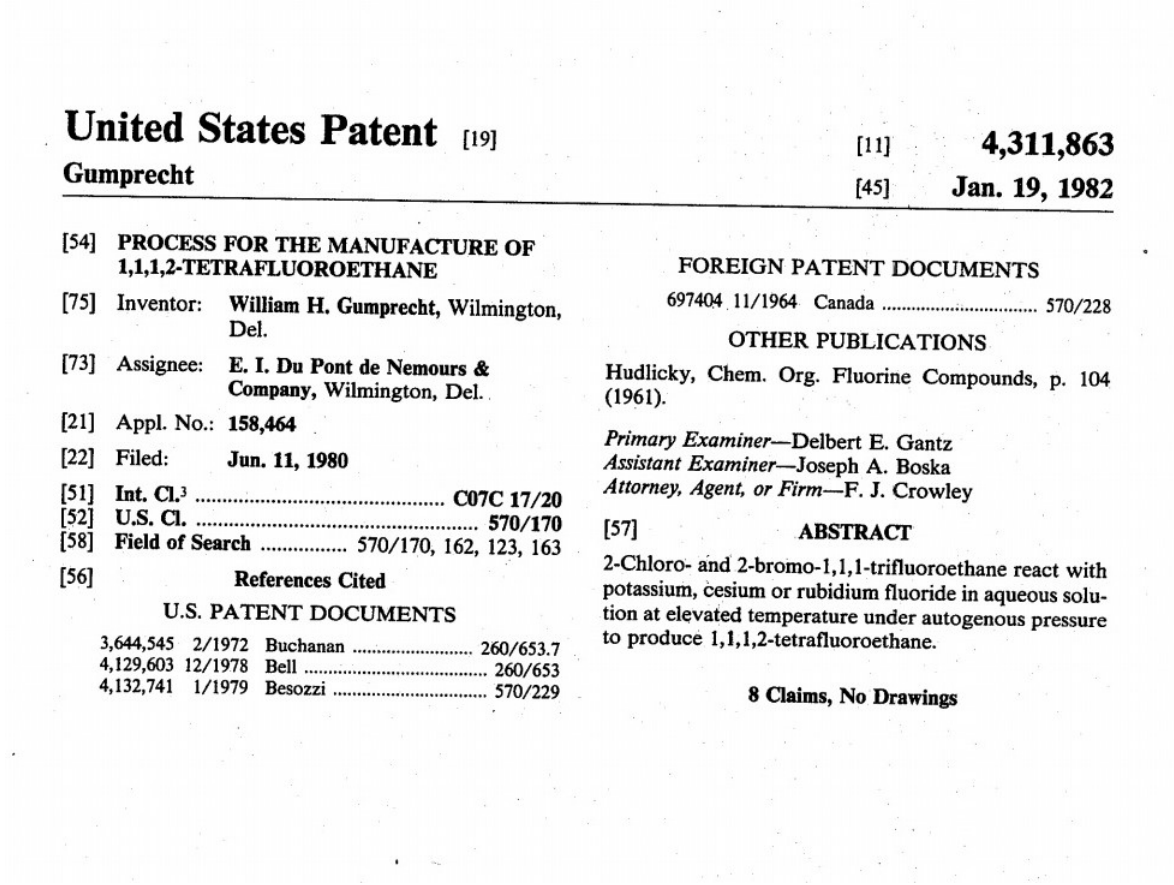
(b) Articles: raw effect (left) and placebo tests (right)

Figure 28: Synthetic control for Annex B compounds using the weak rule (LDA 5 topics, unweighted means)

Notes: Figures 28a and 28b display the results of the synthetic control method for Annex B compounds for articles and patents. In all cases, the method is implemented using the topic proportions of a LDA model with 5 topics and the weak rule for assigning documents to molecule groups. Weighted means of topic proportions are used for patent and unweighted means for articles because these are the specifications that yielded lowest pre-RMPSE. 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, the non-parametric tests to evaluate the significance of the results; black lines show the effect on the treated group relative to the control group, while each gray line is a placebo test performed on an unit drawn from the donor pool.

8 Others

- Figure 29 is an example of a patent granted to DuPont in 1982.
- Figure 30 is an example of a patent granted to DuPont in 1998.



United States Patent [19]

Gumprecht

[11] **4,311,863**

[45] **Jan. 19, 1982**

[54] **PROCESS FOR THE MANUFACTURE OF 1,1,1,2-TETRAFLUOROETHANE**

[75] Inventor: **William H. Gumprecht**, Wilmington, Del.

[73] Assignee: **E. I. Du Pont de Nemours & Company**, Wilmington, Del.

[21] Appl. No.: **158,464**

[22] Filed: **Jun. 11, 1980**

[51] **Int. Cl.**[3] **C07C 17/20**

[52] **U.S. Cl.** **570/170**

[58] **Field of Search** 570/170, 162, 123, 163

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,644,545	2/1972	Buchanan	260/653.7
4,129,603	12/1978	Bell	260/653
4,132,741	1/1979	Besozzi	570/229

FOREIGN PATENT DOCUMENTS

697404	11/1964	Canada	570/228

OTHER PUBLICATIONS

Hudlicky, Chem. Org. Fluorine Compounds, p. 104 (1961).

Primary Examiner—Delbert E. Gantz
Assistant Examiner—Joseph A. Boska
Attorney, Agent, or Firm—F. J. Crowley

[57] **ABSTRACT**

2-Chloro- and 2-bromo-1,1,1-trifluoroethane react with potassium, cesium or rubidium fluoride in aqueous solution at elevated temperature under autogenous pressure to produce 1,1,1,2-tetrafluoroethane.

8 Claims, No Drawings

Figure 29: Snapshot of a patent

References

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

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

US005709092A

United States Patent [19]

Shiflett

[11] **Patent Number: 5,709,092**

[45] **Date of Patent: *Jan. 20, 1998**

[54] **REFRIGATION PROCESS USING CONSTANT BOILING COMPOSITIONS OF HFC-32, HFC-125 AND HFC-134A**

[75] Inventor: **Mark Brandon Shiflett**, Newark, Del.

[73] Assignee: **E. I. Du Pont de Nemours and Company**

[*] Notice: The term of this patent shall not extend beyond the expiration date of Pat. No. 5,185,094.

[21] Appl. No.: **667,107**

[22] Filed: **Jun. 20, 1996**

Related U.S. Application Data

[63] Continuation of Ser. No. 392,281, Feb. 22, 1995, Pat. No. 5,643,492, which is a continuation of Ser. No. 128,435, Sep. 30, 1993, abandoned, which is a continuation-in-part of Ser. No. 931,371, Aug. 18, 1992, abandoned, which is a continuation-in-part of Ser. No. 649,356, Feb. 1, 1991, Pat. No. 5,185,094, which is a continuation-in-part of Ser. No. 628,000, Dec. 17, 1990, abandoned.

[51] **Int. Cl.**^{6} **C09K 5/04**

[52] **U.S. Cl.** **62/114**; 252/67

[58] **Field of Search** 252/67; 62/114; 510/410

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,810,403	3/1989	Bivens et al.	252/67
4,971,712	11/1990	Gorski et al.	252/68
5,185,094	2/1993	Shiflett	252/67
5,370,811	12/1994	Yoshida et al.	252/67
5,438,849	8/1995	Yoshida et al.	62/498

FOREIGN PATENT DOCUMENTS

0430169	6/1991	European Pat. Off. .

Primary Examiner—Christine Skane

[57] **ABSTRACT**

Ternary mixtures of pentafluoroethane, difluoromethane and tetrafluoroethane are useful as refrigerants, aerosol propellants, heat transfer media, gaseous dielectrics, fire extinguishing agents, expansion agents for polyolefins and polyurethanes, and as power cycle working fluids.

1 Claim, No Drawings

Figure 30: Snapshot of a patent