

# Journal Pre-proof

Have international pollution protocols made a difference?

Elisabeth Thuestad Isaksen

PII: S0095-0696(20)30081-4

DOI: <https://doi.org/10.1016/j.jeem.2020.102358>

Reference: YJEEM 102358

To appear in: *Journal of Environmental Economics and Management*

Received Date: 22 April 2019

Revised Date: 21 May 2020

Accepted Date: 24 June 2020

Please cite this article as: Isaksen, E.T., Have international pollution protocols made a difference?, *Journal of Environmental Economics and Management* (2020), doi: <https://doi.org/10.1016/j.jeem.2020.102358>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc.



# Have international pollution protocols made a difference?\*

Elisabeth Thuestad Isaksen<sup>†‡</sup>

May 21, 2020

## Abstract

Evaluating the effectiveness of international agreements is inherently difficult due to problems such as self-selection, spillovers, anticipation effects, and aggregate-level data. In this paper, I provide new and arguably more credible estimates on the effects of three major pollution protocols on SO<sub>2</sub>, NO<sub>x</sub>, and VOC emissions. I do so by combining a newly available global dataset on emissions dating back to 1970 with a generalized version of the synthetic control method. By constructing “synthetic” controls that mimic the pre-treatment development of each affected country, I mitigate bias caused by self-selection and non-parallel emission trends. The broader data coverage - both geographically and over time - allows me to examine the importance of spillovers and anticipation effects. Results from the estimation show that all three protocols induced emissions reductions well beyond a (synthetic) counterfactual development.

**Keywords:** international environmental agreements, pollution, emissions, synthetic control method

**JEL codes:** Q53, Q58, F53

---

\*I want to thank Torben Mideksa, Andreas Lange, Mari Rege, Kjell Arne Brekke, Jo Thori Lind, Daire McCoy, and audiences at Columbia University, University of Oslo, and WCERE 2018 for helpful comments and suggestions. This research was supported with funding from the Grantham Foundation (through the Grantham Research Institute on Climate Change and the Environment), and the Research Council of Norway.

<sup>†</sup>The Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, 0349 Oslo, Norway. Email: elisabeth.isaksen@gmail.com

<sup>‡</sup>Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, United Kingdom (Visiting Associate).

# Have international pollution protocols made a difference?

May 21, 2020

## Abstract

Evaluating the effectiveness of international agreements is inherently difficult due to problems such as self-selection, spillovers, anticipation effects, and aggregate-level data. In this paper, I provide new and arguably more credible estimates on the effects of three major pollution protocols on  $\text{SO}_2$ ,  $\text{NO}_x$ , and VOC emissions. I do so by combining a newly available global dataset on emissions dating back to 1970 with a generalized version of the synthetic control method. By constructing “synthetic” controls that mimic the pre-treatment development of each affected country, I mitigate bias caused by self-selection and non-parallel emission trends. The broader data coverage - both geographically and over time - allows me to examine the importance of spillovers and anticipation effects. Results from the estimation show that all three protocols induced emissions reductions well beyond a (synthetic) counterfactual development.

**Keywords:** international environmental agreements, pollution, emissions, synthetic control method

**JEL codes:** Q53, Q58, F53

# 1 Introduction

A wide range of environmental problems are characterized by cross-border externalities. To effectively solve these problems, some form of international cooperation is needed. Air pollution has been a major focus of international environmental agreements, and since the 1970s over 60 multilateral treaties, protocols, and amendments have been put in place to address this issue (Mitchell, 2015). The potential of international agreements to deliver emission reductions is intensively discussed in the economic literature, and the majority of theoretical studies postulate that free-riding incentives will undermine the effectiveness of such voluntary efforts.<sup>1</sup> Empirically validating these predictions, however, is methodologically challenging, and to date there are few studies applying a credible framework for causal inference. The fundamental problem is establishing a credible counterfactual for countries voluntarily entering into agreements. How would emissions have evolved in absence of participation? Answering this question is complicated by problems such as self-selection bias, spillovers, and anticipation effects.

In this paper, I examine the 1979 Convention on Long-range Transboundary Air Pollution (LRTAP in the following) and three subsequent protocols with the aim of identifying *causal* effects of the protocols on emissions. The LRTAP framework was the first attempt to deal with problems of air pollution on a broad regional basis, covering countries in Europe and North-America. It was initially conceived as a flexible framework for cooperation, but has later been extended by several protocols containing legally binding targets for emissions reductions. This paper focuses on the first three protocols: the 1985 Helsinki protocol on SO<sub>2</sub>, the 1988 Sofia protocol on NO<sub>x</sub>, and the 1991 Geneva protocol on VOCs.<sup>2</sup> Specifically, I ask the following question: to what extent have ratification of these protocols induced emission reductions beyond what would have happened in absence of ratification?

While there are no empirical examinations of the Geneva protocol as far as I am aware, recent studies find no effect of the Helsinki protocol on SO<sub>2</sub> emissions (e.g., Ringquist and Kostadinova, 2005; Naughton, 2010; Aakvik and Tjøtta, 2011), and significant, but small reduc-

<sup>1</sup>The theoretical literature on international environmental agreements is vast. See e.g., Barrett (1994) and Hoel (1992) for seminal papers, and Benchechroun and Long (2012) or Jørgensen et al. (2010) for literature reviews.

<sup>2</sup>SO<sub>2</sub> is short for sulfur dioxides, NO<sub>x</sub> is short for nitrogen oxides and VOCs is short for volatile organic compounds. All pollutants are associated with adverse health effects, and can travel long distances before depositing and causing damage to ecosystems such as forests and lakes. See Appendix A.2 for more details.

tions in  $\text{NO}_x$  emissions induced by the Sofia protocol (Bratberg et al., 2005; Naughton, 2010).<sup>3</sup> Causal interpretation of these findings rely on several identifying assumptions that I argue are not adequately addressed in the papers – in large part due to data limitations and methodological choices. Specifically, the development in pre-treatment emissions is likely to correlate with participation, treatment effects may spill to nearby control countries, and anticipation effects might materialize before the formal ratification of the protocol.

The aim of this paper is to provide new and arguably more credible causal evidence on the effects of the Helsinki, Sofia, and Geneva protocols on emissions, by combining a new global dataset on emissions with a generalized version of the synthetic control method. The synthetic control method is a data-driven procedure to construct a suitable counterfactual in cases with few treated units (Abadie and Gardeazabal, 2003; Abadie et al., 2010). The synthetic control is constructed by assigning weights to plausibly unaffected countries, where the weights are chosen on the basis of how well the synthetic control approximates the development in important pre-treatment variables, such as past emissions. Here, I generalize the method to multiple treated units by building on recent methodological developments (Dube and Zipperer, 2015; Gobillon and Magnac, 2016). Further, as the method requires data on a sufficiently long pre-intervention period and a large *donor pool* of potential control countries, I apply a new database on  $\text{SO}_2$ ,  $\text{NO}_x$  and VOC emissions covering all countries in the world for the period 1970-2008 (JRC, 2012).<sup>4</sup>

Combining the synthetic control method with the global database allows me to address several shortcomings in the previous literature. First, examining the emission development from 1970 onward reveals that pre-treatment trends are not parallel, hence violating the key identifying assumption underlying any potential outcomes framework. By constructing a unique “synthetic” control unit for each treated country that mimics the pre-treatment trend in emissions and important drivers, I ensure that estimated treatment effects are conditional on a similar pre-treatment trend. Second, as previous studies have almost exclusively relied on a limited sample of European countries participating in the LRTAP framework, they run the risk of underestimat-

<sup>3</sup>See Section 3 for a more comprehensive overview of findings in the previous literature.

<sup>4</sup>The database is constructed by pairing internationally reported activity data with assumptions on sector- and technology-specific emissions factors; see Section 5 and Appendix B for more details.

ing treatment effects if there are positive technology and policy spillovers across countries and protocols.<sup>5</sup> By drawing on a large pool of donor countries to construct the synthetic control, I am able to examine the sensitivity of treatment effects to spillovers by imposing various restrictions on the donor pool. Third, while previous studies have defined the intervention date as the ratification year or the year the protocol entered into force, participating countries may start reducing emissions before the formal implementation of the protocol due to rational expectations, or as a consequence of the dialog leading up to ratification. To capture the full extent of the treatment effect, I follow [Abadie \(2012\)](#) and backdate the intervention date to a time before any anticipation effect can be expected.<sup>6</sup> Lastly, by estimating country-specific as well as average treatment effects, I am able to unveil potential heterogeneous effects across ratifying countries.

While the goal of this paper is to improve on past estimates on the effects of pollution protocols, by carefully considering and addressing key identifying assumptions, it is worth reminding the reader that establishing causal inference of large scale interventions is inherently difficult. In an increasingly globalized world, however, international agreements are bound to play an important role also in the future. Applying the best available tools and data might be our best option if we wish to shed light on the effectiveness of international agreements.

Results from the empirical examination show that all three LRTAP protocols induced emissions reductions beyond a (synthetic) counterfactual development. Looking at the Helsinki protocol, I find that emissions were 23% lower five years into the treatment period compared to the synthetic control, and 22% lower ten years after the intervention date. The deviation from the control group hence occurred in the first five years. The large treatment effect of the Helsinki protocol contrasts the null finding in most previous studies. After disentangling potential causes of this discrepancy, I find the way non-parallel trends are dealt with to be the most important explanation (see next paragraph). Examining the Sofia protocol, I find that emissions were 11% lower than the counterfactual after five years, which is comparable to previous findings.<sup>7</sup> After ten years, the corresponding estimate is 18%. For the Geneva protocol, treatment effects after

<sup>5</sup>See Section 3.2 for a more elaborate discussion of this point.

<sup>6</sup>Specifically, I define the intervention date as the baseline year of each protocol. The baseline year for the Helsinki protocol is the year after the LRTAP Convention (1980). The baseline year for the Sofia and Geneva protocols is the year prior to each protocol meeting, i.e., 1987 and 1990, respectively. I then perform various sensitivity checks to shed light on the importance of the choice of intervention year.

<sup>7</sup>[Bratberg et al. \(2005\)](#) estimate an average, annual treatment effect of around -2.1%.

five and ten years are 15% and 21%, respectively. Using a rank-based inference procedure, I find that the pooled treatment effect of each protocol is statistically significant at a 1% level. The synthetic controls mimic the pre-treatment development in emissions relatively closely, and estimated treatment effects are robust to several adjustments to the predictor set and donor pool. Overall, the results suggest that international agreements have been successful in reducing emissions beyond what they would have been in absence of the interventions. I also find that the protocol targets were more ambitious than a “business as usual” scenario (represented by the synthetic control), and that countries on average either met or exceeded the protocol targets.

Beyond providing empirical evidence on the effects of three major pollution protocols, the empirical examination sheds light on two important methodological issues. First, redefining the intervention date to the year the protocol was ratified or entered into force tend to lower treatment effects. This is particularly the case for the Helsinki protocol on SO<sub>2</sub>. This sensitivity check highlights the importance of carefully considering the de facto intervention date, and to account for anticipation effects in cases where these are likely to occur in order to capture the full extent of the treatment.<sup>8</sup> Second, by contrasting the findings in this paper to a previous study using a difference-in-difference (DiD) set-up with country-specific time trends, I illustrate an important shortcoming of traditional ways of dealing with non-parallel trends. Specifically, I find that their choice of treatment date (median ratification year) and control group (LRTAP countries only) explain some of the discrepancy between results, but the main reason for their small and insignificant treatment effect is due to the inclusion of linear or quadratic country-specific time trends. While their motivation for including such trends is to address violation of the parallel trends assumption, the imposed trends seem to absorb most of the treatment effect. The synthetic control method offers an alternative way of controlling for different trends that avoids the risk of absorbing treatment effects, and can be seen as an extension of the DiD framework to account for time-varying confounders.

The remainder of the paper is structured as follows. Section 2 gives the historical back-

---

<sup>8</sup>An alternative interpretation is that countries experiencing a decline in emissions were more inclined to ratify the protocol. For the Helsinki protocol, however, frequent meetings and negotiations in the years leading up to ratification suggest that anticipation effects were likely; see Section 3.2 and Section 6.5.6 for a more thorough discussion. Further, given we are interested in estimating the effect of the LRTAP framework, the natural intervention date would be the time of the first Convention.

ground of the different protocols. Section 3 reviews the relevant literature and clarifies the contribution of this paper in more detail. Section 4 presents the methodology, while Section 5 describes the data. Section 6 presents the results, and Section 7 concludes.

## 2 Background

### 2.1 The Convention on Long-range Transboundary Air Pollution

In the 1960s scientists started to unravel the link between SO<sub>2</sub> emissions in continental Europe and the acidification of Scandinavian lakes. While the environmental damages were first noted in the 1920s, the idea that air pollutants could travel thousands of kilometers before depositing and creating damage to lakes, rivers and forest did not receive notable attention until the 1960s (UNECE, 2015). The 1972 United Nations Conference on the Human Environment in Stockholm signaled the start of an international initiative to combat transboundary pollution. While several countries remained skeptical of the proclaimed relationship between transboundary pollution and the environmental damages in Scandinavia, new studies in the period 1972-1977 confirmed the hypothesis, which led to a broader scientific consensus (UNECE, 2015).

Having recognized the severity of the problem, and thereby the need for international cooperation, a high-level meeting of the UN Economic Commission for Europe on the Protection of the Environment was held in November 1979 in Geneva. The meeting is formally known as *the Convention on Long-range Transboundary Air Pollution* (LRTAP). Article two of the LRTAP Convention states that “The Contracting Parties (...) shall endeavour to limit and, as far as possible, gradually reduce and prevent air pollution including long-range transboundary air pollution.”.<sup>9</sup> The 1979 Convention was largely a framework agreement, formulating general principles for cooperation on air pollution abatement. It has later been extended by eight specific protocols containing legally binding targets for emission reductions.<sup>10</sup> Table I lists the six subsequent protocols targeting SO<sub>2</sub>, NO<sub>x</sub> or VOC emissions, while Table II lists the country-

<sup>9</sup>The Convention text is available here: <http://www.unece.org/fileadmin/DAM/env/lrtap/full%20text/1979.CLRTAP.e.pdf>.

<sup>10</sup>A convention is a formal agreement between states, and is synonymous with the term treaty. The term protocol is used for an additional legal instrument that complements and adds to a treaty. A protocol is optional because it is not automatically binding for States that have ratified the initial treaty; States must independently ratify a protocol.



specific ratification year for the LRTAP Convention as well as the three protocols which are the focus of this paper, i.e., the Helsinki, Sofia, and Geneva protocols.

**Table I:** International conventions and selected protocols part of the LRTAP framework

Short name	Category	Pollutant(s)	Open for signature	Entry into force	Baseline year(s)
LRTAP	Convention		Nov 1979	Mar 1983	
EMEP	Protocol		Sep 1984	Jan 1988	
Helsinki	Protocol	SO <sub>2</sub>	Jul 1985	Sep 1987	1980
Sofia	Protocol	NO <sub>x</sub>	Oct 1988	Feb 1991	1987
Geneva	Protocol	VOCs	Nov 1991	Sep 1997	1984-1990
Oslo	Protocol	SO <sub>2</sub>	Jun 1994	Aug 1998	
Gothenburg	Protocol	SO <sub>2</sub> , NO <sub>x</sub> , VOCs	Nov 1999	May 2005	

*Notes:* Table lists all LRTAP protocols relevant for SO<sub>2</sub>, NO<sub>x</sub> or VOC emissions. The focus of this paper is on the first three protocols with binding emission reduction targets, i.e., the Helsinki, Sofia, and Geneva protocols. The Oslo and Gothenburg protocols are not analyzed in this paper, but are used to define the end year of the treatment window for the three former protocols. In addition to the six protocols listed in the table, the LRTAP framework also includes two protocols addressing persistent organic pollutants (POPs) and heavy metals: the 1998 *Aarhus Protocol on Persistent Organic Pollutants (POPs)* and the 1998 *Aarhus Protocol on Heavy Metals*.

## 2.2 The LRTAP protocols

The first protocol part of the LRTAP framework was the 1984 *Geneva Protocol on Long-term Financing of the Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe* (EMEP). The protocol did not set any emission reduction targets, but provided a financing scheme to fund future activities and provide information on emissions, transport, and deposition of air pollution. In that way the protocol represented the backbone of the Convention.

The first protocol to contain specific emission reduction targets was the 1985 *Helsinki Protocol on the Reduction of Sulphur Emissions or their Transboundary Fluxes by at least 30 per cent* (the Helsinki protocol in the following).<sup>11</sup> SO<sub>2</sub> emissions had already been established as an important source of acidification of rivers and lakes, and was therefore a natural starting point for the first international protocol. The Helsinki protocol opened for signature in July 1985, and entered into force in September 1987. The protocol committed ratifiers to reduce SO<sub>2</sub> emissions by at least 30% compared to 1980 levels, as soon as possible or by 1993.

<sup>11</sup>The protocol text is available here: <http://www.unece.org/fileadmin/DAM/env/documents/2012/EB/1985.Sulphur.e.pdf>

As more scientific evidence was provided, it became clear that other pollutants, like NO<sub>x</sub>, were also contributing to acidification, and had to be addressed within the international framework. This led to the 1988 *Sofia Protocol concerning the Control of Emissions of Nitrogen Oxides or their Transboundary Fluxes* (the Sofia protocol in the following).<sup>12</sup> The protocol required countries to introduce pollution control measures for the largest existing stationary sources, and to apply national emission standards to major new stationary and mobile sources. The aim stated in the protocol was to reduce NO<sub>x</sub> emissions to 1987 levels by December 1994.<sup>13</sup>

In subsequent years, countries recognized that volatile organic compounds (VOCs), in addition to NO<sub>x</sub>, were contributing to the formation of ground-level ozone and other photochemical oxidant products, causing damage to vegetation and crops. To reduce VOCs, countries adopted the 1991 *Geneva Protocol concerning the Control of Emissions of Volatile Organic Compounds or their Transboundary Fluxes* (the Geneva protocol in the following).<sup>14</sup> Under the Geneva protocol, countries had the opportunity to choose between three different emission reduction targets: a 30 % reduction by 1999 (using a year between 1984 and 1990 as the benchmark)<sup>15</sup>, a 30 % reduction by 1999 within a so-called Tropospheric Ozone Management Area and ensuring that 1999 emissions did not exceed 1988 levels<sup>16</sup>, or a stabilization of emission by 1999 at the same levels as in 1988 - given the 1988 levels did not exceed a specified threshold.<sup>17</sup>

The Helsinki protocol was replaced by the 1994 *Oslo Protocol on Further Reduction of Sulphur Emissions* (the Oslo protocol in the following). While previous protocols roughly prescribed the same percentage emission reductions for all countries, the Oslo protocol derived required emission reductions from cost-effectiveness and effect-based principles.<sup>18</sup> The Oslo,

<sup>12</sup>The protocol text is available here: <http://www.unece.org/fileadmin/DAM/env/lrtap/full%20text/1988.NOX.e.pdf>.

<sup>13</sup>The reference year was 1987 for all countries except the United States, which used 1978 as the reference year.

<sup>14</sup>The protocol text is available here: <http://www.unece.org/fileadmin/DAM/env/lrtap/full%20text/1991.VOC.e.pdf>.

<sup>15</sup>This option was chosen by Austria, Belgium, Estonia, Finland, France, Germany, Netherlands, Portugal, Spain, Sweden, and the United Kingdom (with 1988 as base year), by Denmark (with 1985 as base year), by Liechtenstein, Switzerland and the United States (with 1984 as base year), and by Czech Republic, Italy, Luxembourg, Monaco and Slovakia (with 1990 as base year). Source: [https://www.unece.org/fileadmin/DAM/env/lrtap/vola\\_h1.htm](https://www.unece.org/fileadmin/DAM/env/lrtap/vola_h1.htm)

<sup>16</sup>This option was chosen by Norway (with 1989 as the benchmark year) and Canada (with 1988 as the benchmark year). See Annex I to the Protocol for a definition of a Tropospheric Ozone Management Area.

<sup>17</sup>This option was chosen by Bulgaria, Greece and Hungary.

<sup>18</sup>Specifically, each country's required emission reductions were based on the results of a modeled relationship between SO<sub>2</sub> emissions and the exposure of different ecosystems.

**Table II:** Ratification year of the LRTAP Convention and selected protocols, by country

Country name	Country acronym	LRTAP	Helsinki (SO <sub>2</sub> )	Sofia (NO <sub>x</sub> )	Geneva (VOC)
Austria	AUT	1982	1987	1990	1994
Belgium	BEL	1982	1989	2000	2000
Canada	CAN	1981	1985	1991	
Cyprus	CYP	1991		2004	
Denmark	DNK	1982	1986	1993	1996
Finland	FIN	1981	1986	1990	1994
France	FRA	1981	1986	1989	1997
Greece	GRC	1983		1998	
Iceland	ISL	1983			
Ireland	IRL	1982		1994	
Italy	ITA	1982	1990	1992	1995
Luxembourg	LUX	1982	1987	1990	1993
Malta	MLT	1997			
Netherlands	NLD	1982	1986	1989	1993
Norway	NOR	1981	1986	1989	1993
Portugal	PRT	1980			
Spain	ESP	1982		1990	1994
Sweden	SWE	1981	1986	1990	1993
Switzerland	CHE	1983	1987	1990	1994
Turkey	TUR	1983			
United Kingdom	GBR	1982		1990	1994
United States	USA	1981		1989	

*Notes:* Table shows countries that have ratified the LRTAP Convention before 2000, and that are included in the main sample in the analysis. The years indicate the country-specific ratification year of each protocol. Several countries are excluded from the sample based on large structural changes taking place in the period analyzed, such as former USSR countries, former Yugoslavia (incl. Albania), former Czechoslovakia, Bulgaria, Germany, and Poland. Small islands and microstates like Monaco are also excluded. Some countries lack data on pollution and/or GDP, and therefore need to be excluded from the analysis. There are five countries that have signed but not ratified the Geneva protocol: Canada, Greece, Portugal, Ukraine, and the United States. See Appendix Table A.1 for a complete list of ratifying countries, and Section 5 for a description of the exclusion criteria.

Sofia, and Geneva protocols were later replaced by one single protocol: the 1999 *Gothenburg Protocol to Abate Acidification, Eutrophication and Ground-level Ozone* (the Gothenburg Protocol in the following). The protocol was the first multi-pollutant protocol, covering four different pollutants; SO<sub>2</sub>, NO<sub>x</sub>, ammonia (NH<sub>3</sub>), and VOCs. Similar to the Oslo protocol, the Gothenburg protocol used the principle of cost-effectiveness to set national emission caps.

### 3 The effects of the LRTAP protocols

#### 3.1 Previous findings

Over the past decades, several studies have emerged to shed light on the effectiveness of the different LRTAP protocols.<sup>19</sup> In an early study, [Murdoch et al. \(1997\)](#) investigate the effects of the 1985 Helsinki and 1988 Sofia Protocols. Using a spatial lag model with data for 25 European countries over the period 1980-1990, the authors find that the Helsinki protocol has been more effective in reducing emissions than the Sofia protocol.<sup>20</sup> As the study does not account for the non-experimental nature of the data, estimates are likely to suffer from problems such as self-selection and omitted variable bias. The findings hence cannot be interpreted as causal evidence of how emissions would have evolved in absence of treatment. In a subsequent study, [Murdoch et al. \(2003\)](#) focus on SO<sub>2</sub> emissions and use a joint spatial probit and spatial lag equation to estimate both the participation decision and the level of participation in the Helsinki protocol. Using the same dataset as in [Murdoch et al. \(1997\)](#), they find that voluntary cutbacks beyond the emission target gives incentives to free ride. The study does not say anything about the counterfactual, but focuses on the strategic interaction among ratifiers of the protocol.<sup>21</sup>

Focusing on NO<sub>x</sub>, [Bratberg et al. \(2005\)](#) estimate the effects of the 1988 Sofia protocol using a differences-in-differences (DiD) approach. The authors use a sample of 23 European countries for the period 1985-1996 to evaluate the effect, and find evidence that the protocol led to emission reductions slightly greater than what they would have been in absence of the protocol. The yearly reductions in emissions are found to be around 2.1% greater for countries ratifying the Sofia protocol compared to non-ratifiers. In a similar type of set-up, [Ringquist and Kostadinova \(2005\)](#) estimate the effect of the 1985 Helsinki Protocol. Using data on emissions for 19 European countries for the time period 1980-1994, the authors find that while countries

<sup>19</sup>For an overview of previous empirical studies, see e.g., [Houghton and Naughton \(2016\)](#). Here, I focus on findings related to the Helsinki and Sofia protocol. As far as I am aware, there are no previous empirical studies on the 1991 Geneva protocol.

<sup>20</sup>The authors suggest that the stationary sources of SO<sub>2</sub> emissions, together with the substance traveling shorter distances, makes SO<sub>2</sub> somewhat easier to control than NO<sub>x</sub> emissions.

<sup>21</sup>In a closely related study focusing on the 1994 Oslo protocol, [Finus and Tjøtta \(2003\)](#) use a numerical model to test if countries ratifying the protocol reduced SO<sub>2</sub> emissions beyond the numerical calibrated Nash equilibrium. Comparing actual reductions to a simulated Nash equilibrium, they find that the targets for the Oslo protocol are very close to the simulated Nash equilibrium, and the protocol hence provided little emission cuts beyond Nash behavior.

ratifying the Helsinki Protocol experienced significant emission reductions, the protocol itself had no significant effect on emissions. The same conclusion is reached in [Naughton \(2010\)](#). Using a sample of 16 European countries for the time period 1980-2000, [Naughton \(2010\)](#) estimates the effects of the Helsinki, Oslo and Sofia protocols. Applying a two-stage least squares (2SLS) spatial lag model, the author finds no evidence of an effect of the two first protocols, while the Sofia protocol reduced NO<sub>x</sub> emission levels and trend on average.<sup>22</sup>

A common feature of the previous studies on the LRTAP protocols is the use of a small sample consisting of only European countries, as well as the use of a short pre-intervention time period. [Aakvik and Tjøtta \(2011\)](#) take the literature a step forward by exploiting a newly assembled dataset on SO<sub>2</sub> emissions dating back to 1960, and covering in total 30 European countries. Using a DiD approach, they estimate the effect of the 1985 Helsinki and 1994 Oslo Protocols. As they find evidence that the parallel trends assumption is violated, they introduce country-specific linear and quadratic time trends to mitigate the problem. Using 1986-1993 as the treatment window for the Helsinki protocol and 1995-2001 for the Oslo protocol, the authors find no significant effects of the protocols.

### 3.2 Contribution to the literature

The contribution of this paper is to address methodological and data shortcomings of previous literature, by combining a new dataset with a relatively recent methodology. Specifically, the paper addresses five key identifying assumptions, with the aim of providing more credible causal estimates of the LRTAP protocols. These five dimensions are described in detail below.

First, as participation in international protocols is voluntary, there is likely to be a self-selection bias. In particular, countries that are already on a downward-sloping path might be more inclined to join. The presence of *non-parallel trends* are documented in [Aakvik and Tjøtta \(2011\)](#) as well as this paper (Online Appendix D and F).<sup>23</sup> While [Aakvik and Tjøtta \(2011\)](#) aim

<sup>22</sup>[Vollenweider \(2013\)](#) focuses on a more recent LRTAP protocol; the 1999 Gothenburg protocol. Using a sample of 43 European and Eurasian countries from 1995 to 2008, the author finds no effect of the protocol on SO<sub>2</sub> and NO<sub>x</sub>. A methodological weakness of the paper is that it includes self-deposition of territorial emissions as a covariate in the regressions. As this variable is simply a subset of the outcome variable, it will itself be affected by the policy, leading to a so-called “bad control” problem and hence biased estimates ([Angrist and Pischke, 2008](#)).

<sup>23</sup>Appendix Figure D.2 shows that non-parallel trends are more pronounced for the Helsinki and Sofia protocols than the Geneva protocol.

to mitigate the problem by including country-specific trends in the regressions, this approach is likely to absorb parts of the treatment effects as these are now measured as deviations from an imposed linear or quadratic trend. By contrast, I construct a unique synthetic control unit for each treated country that mimics the pre-treatment trend in emissions and important drivers as closely as possible. This ensures that estimated treatment effects are conditional on a similar pre-treatment trend without running the risk of absorbing large parts of the treatment effects.<sup>24</sup>

Second, by exploiting a newly constructed global dataset on emissions, I am better equipped to address potential problems of *spillovers and complementarities* between protocols. In order to recover unbiased estimates, there cannot be *spillovers* to the control group – an assumption that is hard to meet in the case of large scale interventions like multilateral agreements. The first two assumptions also constitute an inherent trade-off as potential control countries that are similar to the treated country, and hence more likely to meet the common trends assumption, may at the same time be more likely to be (indirectly) affected by the intervention. Geographical and political proximity will likely facilitate diffusion of new policies and technological solutions, and if nearby countries are used as controls, it could potentially lead to an underestimation of the treatment effect.<sup>25</sup> Further, as certain abatement measures are complementary across pollutants, like switching fuels or enhancing energy efficiency, a protocol targeting SO<sub>2</sub> could also have an effect on NO<sub>x</sub> emissions, and vice versa. If such complementarities are substantial, it could further underestimate effects of international cooperation if countries in the control group have ratified other protocols within the LRTAP framework. As previous studies have primarily relied on a sample of European countries signing the 1979 LRTAP Convention, they might downward-bias treatment effects if favorable spillovers are large. Here, I aim to mitigate such concerns by expanding the sample to non-LRTAP countries. This allows me to run sensitivity checks where I exclude countries that are likely to be indirectly affected by a specific LRTAP protocol, such as non-ratifying countries in close geographical proximity.<sup>26</sup>

<sup>24</sup>The potential outcomes literature has typically addressed non-parallel trends by combining a difference-in-difference (DiD) with country-specific time trends, or by matching on lagged outcome variables. O'Neill et al. (2016) and Powell (2017) conduct a systematic comparison of the synthetic control method to these alternative approaches and find the synthetic control method to be the least biased estimator.

<sup>25</sup>Treatment effects could also be overestimated if negative spillovers, such as emissions leakage, dominate.

<sup>26</sup>In the main estimation, I keep countries that have both signed and ratified a specific protocol (e.g., the Helsinki protocol) in the treatment group. The donor pool consists of a trimmed sample of countries that have neither signed nor ratified the protocol in question. The criteria for trimming the donor pool are described in Section 5.

Third, by using a dataset covering a longer pre-treatment period, I am able to account for potential *anticipation effects* by backdating the treatment date. While previous studies tend to use ratification or entry into force as the “intervention” date, countries may start reducing emissions before the formal implementation of the protocol.<sup>27</sup> Such run-up effects are for instance documented by Dekker et al. (2012) in the case of the Helsinki protocol and patenting decisions. If there are signs of anticipation, Abadie (2012) suggests to backdate the intervention to a period before any anticipation effect can be expected in order to capture the full extent of the treatment effect. For the Helsinki protocol on SO<sub>2</sub>, the choice of intervention date is particularly challenging as the focus of the 1979 LRTAP Convention was to combat SO<sub>2</sub> emissions. The first meetings related to the Helsinki protocol were also held in 1983, where the 30% emission reduction target relative to 1980 was proposed (Sliggers et al., 2004). Given that countries had knowledge of the most likely baseline several years prior to the Helsinki protocol meeting in 1985, we might expect to see effects materializing in the years leading up to the meeting. To account for potential anticipation effects, I therefore define the intervention date as the baseline year of the respective protocols. For the Sofia and Geneva protocols, the baseline year corresponds to the year before the protocol meetings. To examine the importance of the choice of intervention date, I do a systematic comparison of treatment effects using four different definitions of the de facto treatment year.

Fourth, massive structural changes took place in Eastern Europe in the period analyzed, such as the fall of the Soviet Union and the reunification of Germany, potentially *confounding* the estimated treatment effects. In contrast to previous studies, I account for the large structural changes in Europe in the period analyzed by excluding all countries heavily affected by the collapse of the Soviet Union, such as former USSR-countries, former Yugoslavia, and Germany.

Lastly, average effects might conceal substantial *heterogeneity*. By applying the synthetic control method, I can estimate country-specific treatment effects and thereby identify which countries increased or decreased emissions. To summarize results, I pool country estimates to arrive at an average, protocol-specific treatment effect, where I build on recent methodology to conduct inference; see next section.

---

<sup>27</sup>See Appendix Figure D.2 for the development in emissions in the period 1970 – 1995/1999 by treatment group and the unweighted donor pool for each of the three protocols.



## 4 Methodology

In this paper, I set out to estimate causal effects of the LRTAP protocols on emissions of SO<sub>2</sub>, NO<sub>x</sub> and VOCs. To address the problem of different pre-treatment trends, I apply the synthetic control method, which uses a weighted combination of control countries to construct a “synthetic” counterfactual. The key idea is that a weighted combination of control countries likely serves as a better comparison than any single country alone. The method was first introduced by [Abadie and Gardeazabal \(2003\)](#), and later extended in [Abadie et al. \(2010\)](#), where they estimate the effect of a large tobacco control program in California.<sup>28</sup> Further, I draw on [Dube and Zipperer \(2015\)](#) and [Gobillon and Magnac \(2016\)](#) to conduct inference on the pooled estimate and construct confidence intervals.

### 4.1 The synthetic control estimator

#### 4.1.1 A single treated unit

I start by presenting a framework for the case of a single treated country. Assume that we have data for a sample of  $J + 1$  countries, where  $j = 1$  denotes the “treated” country, i.e., the country affected by the policy intervention, and  $j = 2, \dots, J + 1$  are countries unaffected by the intervention, i.e., the “donor pool”. In our setting, the intervention is participation in an international pollution protocol, and the outcome of interest is emissions of SO<sub>2</sub>, NO<sub>x</sub> or VOCs. Further, assume that the data spans  $T$  periods, where  $T_0$  is the period prior to intervention. Denoting the intervention as  $D$ , the synthetic control approach assumes that the observed outcome,  $Y_{jt}$ , is the effect from the treatment,  $\alpha_{jt}D_{jt}$ , and the counterfactual outcome,  $Y_{jt}^N$ :

$$Y_{jt} = \alpha_{jt}D_{jt} + Y_{jt}^N = \alpha_{jt}D_{jt} + \theta_t Z_j + \lambda_t \mu_j + \delta_t + \varepsilon_{jt}. \quad (1)$$

Here  $Z_j$  is a vector of observed covariates not affected by the intervention,  $\theta_t$  is a vector of unknown parameters,  $\delta_t$  is a common time factor and  $\varepsilon_{jt}$  is the idiosyncratic error term. In

<sup>28</sup>The method has later been applied to a wide range of topics. Examples include the economic impact of natural resource endowment ([Mideksa, 2013](#)), the effect of economic liberalization on GDP ([Billmeier and Nannicini, 2013](#)), impact of catastrophic natural disasters on economic growth ([Cavallo et al., 2013](#)), the effects of the German reunification on economic costs ([Abadie et al., 2015](#)), the economic costs of organized crime ([Pinotti, 2015](#)), and the effects of the Kyoto protocol on CO<sub>2</sub> emissions ([Almer and Winkler, 2017](#)).



a standard difference-in-differences (DiD) framework, both  $Z_j$  and  $\delta_t$  can be accounted for by comparing the difference between the treatment group and the control group before and after the intervention. As long as the covariates  $Z_j$  do not vary over time, and the time trend  $\delta_t$  is common to all countries, the terms will be differenced out in a DiD set-up. What is left, however, is the term  $\lambda_t \mu_j$ . Here  $\lambda_t$  is a vector of unobserved *time-varying* factors and  $\mu_j$  are the unknown factor loadings. If the factor loadings differ across countries, the assumption of parallel trends for the treated and control countries in absence of intervention will likely be violated. However, if we knew the true factor loadings  $\mu_1$  for the treated country, we could construct an unbiased control by using donor states whose factor loadings average to  $\mu_1$ .

The idea of the synthetic control method is to construct a vector of weights  $\mathbf{W}$  over  $J$  donor states such that the weighted combination of donor states closely mimics the outcome of the treated country in the pre-intervention period. This weighted combination of donor units is called the synthetic control. Given a good match, we can difference out the time-varying term  $\lambda_t \mu_j$ . More formally, for the treated country, I define the  $(k \times 1)$  vector of pre-treatment characteristics as  $X_1 = (Z_1', Y_j^{K_1}, \dots, Y_j^{K_L})$ , where  $Y_j^{K_i}$  are  $L$  linear combinations of pre-treatment outcomes. Analogously, I define the  $(k \times J)$  matrix containing the same characteristics for the  $J$  donor countries as  $X_0$ . The synthetic control procedure chooses donor weights  $\mathbf{W}$  to minimize the distance between pre-treatment characteristics  $X_1$  and  $X_0$  of the treated country and untreated countries. More specifically, the method minimizes the mean square prediction error (MSPE) over  $k$  pre-treatment characteristics:

$$\sum_{m=1}^k v_m (X_{1m} - X_{0m} \mathbf{W})^2, \quad (2)$$

where  $v_m$  measures the relative importance of the  $m^{th}$  predictor. Given the optimal weights  $w_j^*$  for each of the  $j = 2, \dots, N$  donors, the synthetic control at any time  $t$  is simply the weighted combination of the outcome variable (i.e., pollution) in the donor countries:  $\sum_{j=2}^N w_j^* Y_{jt}$ .<sup>29</sup> The estimate of the treatment effect  $\alpha_{1t}$  is therefore the difference between pollution in the treated

<sup>29</sup>The weights are non-negative and restricted to sum to one. This implies that synthetic controls are weighted averages of the units in the donor pool. Restricting country weights to sum to one is warranted only if the dependent variable is rescaled, so it is not affected by country size. As described in Section 5.1, I use normalized variables of pollution per capita as the outcome variable, which would warrant such a restriction.

country  $Y_{1t}$  and pollution in the synthetic country  $\sum_j w_j^* Y_{jt}$  at any post-treatment time  $t \geq T_0$ :

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^N w_j^* Y_{jt} \quad (3)$$

In the post-intervention period  $t = T_0, \dots, T$ , the average difference between the treated and synthetic control outcomes is given by

$$\hat{\beta}_1 = \frac{1}{T} \sum_{t=T_0}^T (Y_{1t} - \sum_{j=2}^N w_j^* Y_{jt}) \quad (4)$$

In the analysis, the outcome variable is normalized to 100 in the year prior to treatment (see Section 5.1). This means that we can interpret  $\hat{\beta}_1$  as the average difference in percentage points between the treated and the synthetic counterfactual development.

#### 4.1.2 Multiple treated units: pooled estimate

In the case of the LRTAP protocols, there are multiple treated countries. I therefore generalize the framework to multiple units. Denoting treated countries by subscript  $e$ , where  $e = 1, \dots, E$ , I calculate an annual country-specific treatment effect  $\hat{\alpha}_{e1t}$  and an average country-specific treatment effect  $\hat{\beta}_{e1}$  by using equations 3 and 4. The *pooled* treatment effect can be expressed as:

$$\bar{\alpha}_{e1t} = \frac{1}{E} \sum_{e=1}^E \hat{\alpha}_{e1t} \quad \bar{\beta}_{e1} = \frac{1}{E} \sum_{e=1}^E \hat{\beta}_{e1}, \quad (5)$$

where  $\bar{\alpha}_{e1t}$  is the pooled treatment effect for a given year, and  $\bar{\beta}_{e1}$  is the pooled treatment effect averaged over the post-treatment period. I also calculate an alternative pooled treatment effect, the Hodges-Lehmann (HL) pooled estimate, which I explain in detail in Section 4.3.2.

## 4.2 Statistical inference

### 4.2.1 A single treated unit

To assess the statistical significance of a single country's estimated treatment effect, I use placebo-based inference. This involves running a number of falsification tests, or “placebo tests”, for the countries in the donor pool. The estimated treatment effect for the treated unit is

then compared to the distribution of placebo effects. Specifically, I estimate treatment effects  $\hat{\beta}_j$  for each of the  $j = 2, \dots, N$  donor countries by repeating the procedure described in Section 4.1, but using the remaining  $N - 2$  donor countries. These placebo runs are used to evaluate the statistical significance of the true treatment estimate. In the case of a single treated country, I compare the magnitude of the treatment effect for the treated country to the treatment effects of the placebo runs.<sup>30</sup> I then rank the treatment effects according to magnitude. This allows me to construct a percentile rank statistics  $p$  for the treated country:

$$p_{1t} = \hat{F}(\alpha_{1t}) \quad p_1 = \hat{F}(\beta_1), \quad (6)$$

where  $F$  is the empirical cumulative distribution function (CDF) of the coefficients  $\hat{\alpha}_{jt}$  or  $\hat{\beta}_j$ . As the percentile rank is approximately uniformly distributed, I can determine whether the rank of the treated state,  $p_1$ , lies in the tails of the distribution. Using a two-sided statistical significance level of 5 percent, I reject the null of  $\beta_1 = 0$  when  $p_1 < 0.025$  or  $p_1 > 0.975$ .<sup>31</sup>

#### 4.2.2 Multiple treated units: pooled estimate

To conduct inference on the pooled treatment effect in equation 5, I construct a test statistic  $\bar{p}$  which is the mean of the percentile ranks of treated countries:

$$\bar{p}_t = \frac{1}{E} \sum_{e=1}^E p_{et} \quad \bar{p} = \frac{1}{E} \sum_{e=1}^E p_e \quad (7)$$

If we assume that ranks are independent across treated countries, the exact distribution of  $\bar{p}$  can be calculated using the Irwin-Hall distribution of the sum of  $E$  independent uniform random variables. This procedure is described in detail in Online Appendices C.1 - C.2. Alternatively, we can form a distribution of the mean percentile ranks by randomly permuting the treatment status, see Appendix C.3.<sup>32</sup> The permutation exercise is far more computationally intensive

<sup>30</sup>To account for the fact that a poor pre-treatment fit might give rise to larger post-treatment deviations, I trim the donor pool based on pre-treatment fit. Specifically, I trim the donor pool down to the 42 countries with the lowest mean square prediction error (MSPE).

<sup>31</sup>Note that the number of available donors limits the range of confidence levels I can implement for a single treated event. In order to assess a two-sided 5 percent level of significance, I need at least 41 donor countries.

<sup>32</sup>The permutation procedure has similarities to the procedures described in Section 4.5 in Dube and Zipperer (2015) and in the Results section in Gobillon and Magnac (2016).

than using the Irwin-Hall distribution. Also, the small number of actually observed percentile ranks will influence the cut-off values. I therefore focus on the cut-off values from the Irwin-Hall distribution when evaluating statistical significance (see Appendix Table C.3), but use the cut-off values from the permutation procedure in robustness checks.<sup>33</sup>

### 4.3 Constructing confidence intervals

#### 4.3.1 A single treated unit

In the case of a single treated country, we can invert the percentile ranks,  $p_{1t}$ , to construct confidence sets. Inverting the percentile rank means that I ask for what values of  $\tau$  does the following inequality hold:

$$0.025 \geq \hat{F}_{1t}(\alpha_{1t} - \tau) \geq 0.975 \quad (8)$$

The term  $\hat{F}_{1t}(\alpha_{1t} - \tau)$  is referred to as the *adjusted* country-specific rank,  $p_{1t}(\tau)$ . The 95 percent confidence interval is the set of  $\tau$  not rejected using the critical values 0.025 and 0.975.

#### 4.3.2 Multiple treated units: pooled estimate

To construct confidence intervals for the pooled effect, I invert the mean rank statistic  $\bar{p}_t$ . This means that I ask for what values of  $\tau$  does the following inequality hold:

$$\text{Lower critical value} \geq \frac{1}{E} \sum_{e=1}^E \hat{F}_{e1t}(\alpha_{e1t} - \tau) \geq \text{Upper critical value}, \quad (9)$$

where  $\frac{1}{E} \sum_{e=1}^E \hat{F}_{e1t}(\alpha_{e1t} - \tau)$  is the *mean adjusted rank*,  $\bar{p}_t(\tau)$ .<sup>34</sup> The 95 percent confidence interval for the pooled effect is the set of  $\tau$  such that the mean adjusted rank  $\bar{p}_t(\tau)$  lies within the critical values presented in Appendix Table C.3. By collapsing the pooled confidence intervals, I get the Hodges-Lehmann (HL) point estimate (Hodges Jr and Lehmann, 1963). The HL estimate is simply the mean of the upper and lower confidence bounds. In the case of a single treated country, the mean and the HL point estimate are the same. In the case of multiple

<sup>33</sup> Additionally, I also address the potential problem of rank dependency by performing a randomization procedure that constrains the permutation of treatment status by forcing the “treated” countries to be located geographically close to each other. The procedure is described in detail in Appendix C.4, while robustness checks with these alternative cut-off values are presented in Section 6.5.

<sup>34</sup> A similar type of procedure is described in Gobillon and Magnac (2016).

treated countries, the mean and the HL point estimate are not necessarily the same. If outlying estimates of individual treatment effects heavily influence the mean estimate, the mean and the HL estimate will differ substantially. While the mean estimate has a more clear interpretation, the HL estimate is more robust to outliers (Gobillon and Magnac, 2016).

#### 4.4 Requirements and caveats

Before applying the synthetic control method, it is important to have a clear understanding of the conditions that may influence the suitability of the method as a tool for policy evaluation.<sup>35</sup> First, if the outcome variable is *highly volatile*, the synthetic control method may not be able to distinguish a treatment effect from random shocks to the outcome variable. In particular, if the magnitude of impacts from an intervention is similar to the volatility of the outcome variable, treatment effects are difficult to detect.<sup>36</sup> Second, if potential control countries adopt a *similar type of intervention* as the one adopted by the treated country, they should not be included in the donor pool.<sup>37</sup> It is also important to eliminate from the donor pool any country that may have suffered *large idiosyncratic shocks* to the outcome of interest during the period analyzed. Third, the *differences in the characteristics* of the affected country and the synthetic control should not be too big. Also, if a country had particularly low or particularly high levels of emissions before the treatment date relative to the countries in the donor pool, then no weighted average of countries in the donor pool will be able to closely reproduce the pre-intervention emissions for the country. As a way around this, Abadie (2012) suggests to transform the outcome to time differences or growth rates.<sup>38</sup> Fourth, while countries in the donor pool should not be too different from the treated countries, they should at the same time be unaffected by the intervention. If *spillover effects* are likely to be substantial, it may be advisable to exclude countries expected to be indirectly affected. There is hence a tension between the issue of no spillovers and having comparable countries in the donor pool. Fifth, the synthetic control

<sup>35</sup>Most of the conditions are also relevant to other policy evaluation tools, including difference-in-differences.

<sup>36</sup>This problem arises if the volatility is intrinsic to the treated country. Common shocks affecting all other countries can be differentiated out by choosing a suitable synthetic control.

<sup>37</sup>As an example, Abadie et al. (2010) discard from the donor pool several states that adopted large-scale tobacco programs during the sample period of the study.

<sup>38</sup>The same logic is used in a DiD framework; even if the level of the outcome variable cannot be reproduced, there are cases when a control group can reproduce the *changes* in the outcome variable for the treatment group.

estimator may be biased if forward looking countries *react in advance of the policy intervention*, or if certain components of the intervention are put in place before the formal implementation. If there are signs of anticipation, [Abadie \(2012\)](#) recommends to backdate the intervention to a period before any anticipation effect can be expected.

## 5 Data and descriptives

The following subsections describe the data sources used in the analysis. The choice of outcome variables, predictors, treated and donor countries, and treatment window are based on the recommendations and caveats discussed in Section 4.4.<sup>39</sup>

### 5.1 Outcome variables

The outcome variables of interest are SO<sub>2</sub>, NO<sub>x</sub> and VOC emissions. While previous studies on the LRTAP protocols have primarily relied upon the officially reported data to the EMEP, this dataset lacks emissions for non-LRTAP countries as well as consecutive data prior to 1985 and is therefore not suitable for our purpose.<sup>40</sup> Instead, I compile country-level emissions of SO<sub>2</sub>, NO<sub>x</sub> and VOC from the Emission Database for Global Atmospheric Research (EDGAR in the following) ([Janssens-Maenhout et al., 2011](#)).<sup>41</sup> EDGAR provides emissions of air pollutants by country and sector for nearly all countries in the world over the period 1970-2008. The database is constructed by pairing internationally reported activity data with assumptions on sector- and technology-specific emissions factors. The bottom-up methodology is applied consistently for all world countries; see Appendix B for further details.<sup>42</sup>

<sup>39</sup>A summary of key adjustments made to the data is presented in Appendix Table D.1.

<sup>40</sup>The EMEP dataset provides emissions data for the years 1980, 1985, and then annually from 1990 and onward for the countries that have signed the 1980 LRTAP Convention. The data set is available at [http://www.ceip.at/ms/ceip\\_home1/ceip\\_home/webdab\\_emepdatabase/reported\\_emissiondata/](http://www.ceip.at/ms/ceip_home1/ceip_home/webdab_emepdatabase/reported_emissiondata/).

<sup>41</sup>The development of EDGAR is a joint project of the European Commission Joint Research Centre and the Netherlands Environmental Assessment Agency. The dataset is available at the following website: <http://data.jrc.ec.europa.eu/dataset/jrc-edgar-emissiontimeseriesv42>.

<sup>42</sup>Note that emissions in the EDGAR database may not necessarily correspond to the officially reported EMEP data, as the EDGAR database relies on a technology based emission factor approach. The same methodology is applied to all countries to ensure comparability. Note also that another dataset on SO<sub>2</sub> emissions ([Stern, 2006](#)), dating back to 1960, has been used in [Aakvik and Tjøtta \(2011\)](#). For consistency reasons, I use the same data source (EDGAR) for all three pollutants.

To enhance comparability across countries, I focus on emissions per capita rather than total emissions.<sup>43</sup> The outcome variable is then normalized by setting emissions per capita to be equal to 100 in the year before the intervention.<sup>44</sup> The normalization allows me to pool estimates to arrive at an average, protocol-specific treatment effect. For the transformation of the outcome variable to make sense, I need to assume that donor countries with lower (higher) pollution levels are able to reproduce *trends* in emissions for treated countries with a higher (lower) pollution level. A similar logic is also used in a difference-in-differences framework, see Section 4.4.<sup>45</sup>

## 5.2 Predictors

The key predictor used in the synthetic control method is pre-intervention values of the outcome variable, which in our case is normalized emissions per capita. By constructing a synthetic control that closely mimics past emission trends of the treated unit, the approach will, in theory, indirectly account for unobserved, time-varying variables that influence emissions. To ensure that emission *levels* are not too different for the treated unit and the synthetic control, I also include the per capita pollution level in the treatment year as a second predictor.

While past emissions is the most important predictor, we also want to make sure that treated and donor countries are relatively similar along dimensions that we think are important for emission trends, such as the state of economic development.<sup>46</sup> To this end, I collect data on GDP per capita (in constant 2005 US\$), GDP growth, the share of GDP from industrial activities, and population growth.<sup>47</sup> I also collect data on the share of fossil fuels of total energy use, as countries with a relatively "dirty" energy mix are likely to have a larger potential for emis-

<sup>43</sup>Population data is compiled from [The World Bank \(2015\)](#).

<sup>44</sup>A similar approach is taken in [Cavallo et al. \(2013\)](#) and [Almer and Winkler \(2017\)](#).

<sup>45</sup>In Appendix J I also estimate treatment effects using total emissions and emissions as a share of GDP as outcome variables. Both approaches yield relatively similar results as using emissions per capita as the outcome.

<sup>46</sup>A relevant concept here is the so-called Environmental Kuznets Curve hypothesis, which postulates an inverted u-shaped relationship between pollution and GDP, see e.g., [Dinda \(2004\)](#). Although the empirical support for the hypothesis is mixed, comparing countries at different stages in the economic development could imply that richer countries are on a downward-sloping path while poorer countries are on an upward-sloping path. At the same time, the synthetic control method is designed to mitigate such problems, by constructing synthetic controls that approximate the development in emissions in the period prior to the intervention.

<sup>47</sup>Industry share of GDP is only available for a subset of countries and years. To ensure that all treated countries and a sufficiently large donor pool is included in the analysis, missing values are replaced by values from the latest available year.

sion reductions. Following previous literature ([Ringquist and Kostadinova, 2005](#); [Vollenweider, 2013](#)), I also account for the stronger incentive for heavily forested countries to participate in pollution protocols by compiling data on the share of land area covered by forest.<sup>48</sup> All data except emissions are compiled from [The World Bank \(2015\)](#).

The data on country characteristics can be used to exclude donor countries with very different pre-treatment characteristics and/or be used as predictors to construct the synthetic controls. Note that including additional predictors beyond past emissions imply several trade-offs. First, some of the country characteristics are missing for the pre-intervention period, or only available for a small sub-sample of countries.<sup>49</sup> Including these variables as predictors hence implies dropping a substantial number of the countries from the analysis, or requires some form of imputation. Second, adding more predictors will necessarily lower the weights assigned to other key predictors, like past emissions, potentially leading to a poorer pre-intervention match on these variables. In the main specification, I have chosen to include the following four predictors: normalized emissions per capita in the years prior to the intervention, emissions per capita in the treatment year, GDP per capita in the treatment year, and the share of fossil fuel of total energy use in the treatment year.<sup>50</sup> In robustness checks, I examine the sensitivity of results to changing the predictor set.

### 5.3 Treated and donor countries

Information on participation in environmental protocols is compiled from the International Environmental Agreements database project ([Mitchell, 2015](#)).<sup>51</sup> In the analysis, I define the treatment group as countries that have both signed and ratified the protocol in question before the year 2000. The donor pool is defined as countries that have neither signed nor ratified the protocol before the year 2000.<sup>52</sup> While the initial data set covers over 170 countries, I make several

<sup>48</sup>Note that forest area data is only available as 5-year averages, where the first available year is 1990.

<sup>49</sup>Appendix Table A.1 shows which LRTAP countries lack data on GDP.

<sup>50</sup>The relative importance of each predictor ( $v_m$  in equation 2) are set to the following values: normalized emissions: 0.5, emission levels: 0.4, GDP: 0.095, and fossil share: 0.005. By fixing the weights  $v_m$ , I ensure that predictors are weighted in the same way for all countries. In robustness checks, I use a data-driven approach to assign country-specific predictor weights.

<sup>51</sup>The database contains information on when a protocol opened up for signature, the date it entered into force, as well as each country's signature and ratification date. The database is available at <http://iea.uoregon.edu/>.

<sup>52</sup>Countries that have only signed but not ratified the protocol in question are dropped from the analysis, as it is not clear whether they should be regarded as treated or donor countries. For the Geneva protocol, there are five



adjustments that substantially lower the number of countries in the sample. First, countries should not experience country-specific structural shocks to the outcome variable that coincide with the intervention. Based on this, I exclude countries heavily affected by the fall of the Soviet Union, such as former USSR countries, former Yugoslavia, former Czechoslovakia, Germany, and Poland, and countries experiencing long-lasting conflicts and wars during the treatment period. Second, as the majority of treated countries are high-income countries, I drop the poorest quintile from the sample.<sup>53</sup> Third, I exclude small islands and microstates, such as Monaco and Lichtenstein, as well as countries with an extremely volatile emission development.<sup>54</sup>

Lastly, as discussed in Sections 3.2 and 4.4, it may be advisable to exclude donor countries that are likely to experience substantial spillovers. Spillover, or indirect effects, can both increase or decrease emissions in countries not covered by the protocol in question. At the one hand, technology and policy diffusion might lead countries not covered by the protocol to reduce their emissions. This might particularly be the case for similar countries in close geographical proximity to treated countries. Also, European countries not covered by the protocol might be affected via new EU directives triggered by the protocol. If this is the case, including these countries in the donor pool will likely underestimate treatment effects. The same is true if there are strong complementarities between abatement measures for the three pollutants, and countries in the donor pool have ratified another LRTAP-protocol. In particular,  $\text{SO}_2$  and  $\text{NO}_x$  are often emitted as co-pollutants and efforts to reduce one of these pollutants will likely effect both (see Appendix A.2). On the other hand, spillovers can also take the form of emission leakage through re-allocation of pollution-intensive industries, or via input markets, which could increase emissions in countries not covered by the protocol. This would instead overestimate the effect of the protocol.

As the emissions dataset covers a wide range of countries both inside and outside the LRTAP

---

countries that have signed but not ratified the protocol: Canada, Greece, Portugal, Ukraine, and the United States. For the Helsinki and Sofia protocol, all signatory countries have also ratified the protocols.

<sup>53</sup>Specifically, I exclude all countries with GDP per capita in the lowest 20th percentile in 1980. In robustness checks, I test the sensitivity of trimming the donor pool by GDP.

<sup>54</sup>I use a criteria where I exclude countries if emissions in the peak year is more than three times higher than the minimum emissions in the period analyzed. I also exclude Norway from the treatment group when analyzing the Geneva protocol due to the drastic fluctuations in VOCs caused by the accelerating oil production from 1975. While storage and transportation of crude oil have large impacts on VOCs due to evaporation of chemicals,  $\text{NO}_x$  and  $\text{SO}_2$  are primarily caused by fuel combustion, see Appendix A.2.

framework, I am able to put different restrictions on the donor pool to see to what extent this influence treatment effects. In the main specification, I have chosen to define the donor pool as all countries that have neither signed nor ratified the specific protocol in question.<sup>55</sup> This implies that I include both European and non-European countries in the donor pool, as well as countries that might have ratified other LRTAP-protocols than the one in question. The main specification leaves us with 43-51 countries in the donor pool and 12-17 treated countries; see Appendix Table D.2 for a list of countries. While including countries outside the LRTAP framework in the donor pool will likely dilute the problem of favorable spillovers, treatment effects might still be underestimated if the synthetic control is heavily based on nearby countries that have ratified other LRTAP protocols. To further explore the role of spillovers, I perform several sensitivity checks that are presented in Section 6.5. In particular, I examine the effect of (i) dropping all LRTAP countries from the donor pool, and (iii) restricting the sample of treated and donor countries to those participating in the LRTAP framework.

## 5.4 Time period

*Intervention year:* Previous studies have typically used each country's ratification year as the intervention date. This might be problematic if there are anticipation effects, or if certain components were in place prior to the formal implementation. In the case of the Helsinki protocol, the first LRTAP meeting in November 1979 represented the start of the international cooperation efforts. The primary focus of the first meeting was SO<sub>2</sub> emissions, and we can therefore expect there to be anticipation effects for this pollutant. Further, the baseline year for the Helsinki protocol was set to 1980. If this was known in advance of the meeting in 1985, it could give an incentive for early reductions as emission targets are stated as reductions relative to the baseline year. To capture the full effect of the Helsinki protocol, I define the intervention year as the baseline year of the protocol (1980). This backdating of the intervention date is in line with the recommendations in Abadie (2012). By the same token, I define the intervention date for the Sofia protocol as 1987 (baseline year and the year before the protocol meeting) and the intervention date for the Geneva protocol as 1990 (the latest baseline year and the year before the

---

<sup>55</sup>An exception is Iceland, which is excluded from the donor pool in the analysis of the Sofia and Geneva protocol based on what is interpreted as a positive policy spillover; see Section 6.2 for details.

protocol meeting). In Section 6.5.6, I examine the sensitivity of the main results by redefining the intervention year to later years.

*Pre and post periods:* The pre period is defined as the years from 1970 until the year before the intervention. The post period is defined as the year of the intervention until the protocol in question was replaced by a new one. The Oslo protocol opened up for signatures in 1994, and was meant to replace the Helsinki protocol. The Gothenburg protocol replaced all previous protocols on SO<sub>2</sub>, NO<sub>x</sub> and VOCs, and opened up for signatures in 1999. This implies the following post periods for the three protocols: 1980-1994 for the Helsinki protocol, 1987-1999 for the Sofia protocol, and 1990-1999 for the Geneva protocol.<sup>56</sup>

## 5.5 Descriptives

**Table III:** Summary statistics by treated and donor countries. 1980.

	Treated countries					Donor pool				
	mean	(sd)	min	max	N	mean	(sd)	min	max	N
<b>Outcome variables:</b>										
SO <sub>2</sub> per capita (1980)	74	(33)	34	152	12	35	(38)	2	192	51
NO <sub>x</sub> per capita (1987)	43	(21)	23	93	17	22	(23)	4	112	43
VOC per capita (1990)	51	(14)	36	89	12	43	(42)	7	238	44
<b>Predictors:</b>										
GDP per capita (2005 USD)	27,788	(6,609)	20,991	42,657	12	8,911	(13,216)	633	81,947	51
Fossil energy use (% of total)	82	(14)	58	98	12	74	(26)	16	100	51
GDP growth (%)	2.36	(1.73)	-0.49	5.39	12	3.33	(6.59)	-13.23	23.87	51
Population growth (%)	0.38	(0.35)	0.00	1.29	12	2.37	(1.52)	0.12	9.15	51
Forest area (% of land area)	35	(19)	10	72	12	30	(23)	0	73	51
Industry GDP (% of total)	29	(4)	19	34	12	32	(13)	12	73	51

*Notes:* Table shows means, standard deviations (sd), minimum values, maximum values, and the number of countries (N). Unless stated otherwise, data is from 1980. Emissions are given in kilogram (kg) per capita. GDP is given in constant 2005 USD per capita. SO<sub>2</sub> per capita and all predictors are based on the sample used for estimating effects of the Helsinki protocol. For NO<sub>x</sub> per capita, the sample corresponds to the one used to estimate effects of the Sofia protocol. For VOCs per capita, the sample corresponds to the one used to estimate effects of the Geneva protocol. Forest area data is from 1990 (first available year) and reflect a 5-year average.

Table III presents summary statistics by treated and donor countries.<sup>57</sup> Although treated countries have on average higher emission levels per capita than countries in the donor pool, there is common support for all three pollutants. In other words; the minimum and maximum

<sup>56</sup>Alternative approaches to define the post period could be to (i) use the end year of the emission targets (i.e., 1993 for the Helsinki protocol, 1994 for the Sofia protocol, and 1999 for the Geneva protocol), or (ii) use the year the Oslo and Gothenburg protocols entered into force (i.e., 1998 and 2005, respectively).

<sup>57</sup>Appendix Table D.2 lists the countries assigned to the treatment group and donor pool in the three protocols. As some of the restricting criteria presented above are pollutant-specific, the donor pool will be slightly different for the three pollutants. Further, as the treatment group is defined as countries signing and ratifying the protocol in question, the number of treated countries will also vary.

values for the treatment group are within the interval of minimum and maximum values for the donor pool. This implies that it should be possible to construct a synthetic control that closely matches the emission level of each treated country. The same is true for all country characteristics except population growth. However, as the outcome variable is measured in per capita terms, population growth is indirectly accounted for. While there is common support for most variables, Appendix Figure D.1 shows that there are relatively few donor countries in the upper parts of the income distribution. This means that it will be challenging to construct a synthetic control for the richest countries in the treatment group (e.g., Switzerland and Norway) that closely reproduces GDP per capita, as there are fewer donor countries to choose among.

## 6 Results

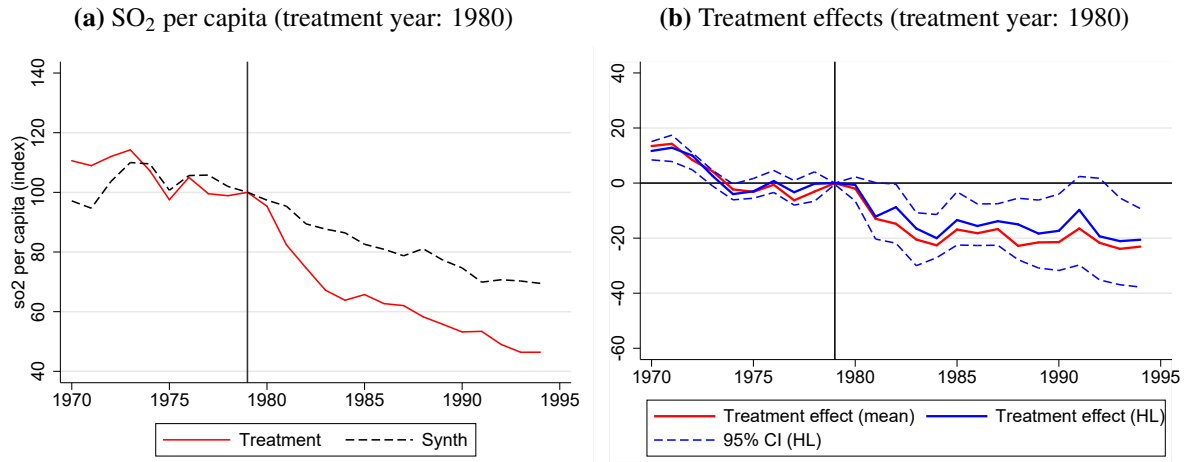
In the following, I report results from the synthetic control method, where countries that have signed and ratified the protocol in question are defined as the treatment group. The outcome variables are normalized values of SO<sub>2</sub>, NO<sub>x</sub> and VOC emissions per capita.

### 6.1 Effects of the Helsinki protocol on SO<sub>2</sub> emissions

Figure 1 and Table IV summarize the effects of the Helsinki protocol on SO<sub>2</sub> emissions.<sup>58</sup> Using the baseline year as the intervention date (1980), countries that ratified the Helsinki protocol experienced a sharp decline in emissions in the post-treatment period, see Figure 1a. The rate of decline is larger for the treated countries than for the synthetic control. From Figure 1b we observe that treatment effects are significantly different from zero in the post-treatment period. On average, SO<sub>2</sub> emissions reductions were 17-18% larger compared to the synthetic control in the treatment period (see Table IV) and the effect is significant at a 1% level using a two-sided test. Looking at the development over time, most of the reduction materializes in the first 5 years. After 5 years, emissions are 20-23% lower than the synthetic control, and emissions fluctuate around this level for the rest of the period.

---

<sup>58</sup>See Appendix Table D.2 for a list of the treated and donor countries.

**Figure 1:** Effects of the Helsinki protocol on SO<sub>2</sub> emissions

Notes: Panel (a) shows the development in emissions for the treatment group (red, solid line) and the synthetic control group (black, dashed line). Emissions in the year before treatment is normalized to 100. Panel (b) shows yearly treatment effects. The solid red line corresponds to the average, yearly treatment effects  $\hat{\alpha}_{e1t}$  estimated from equation 5. The solid blue line indicates the HL point estimate (see Section 4.3.2 for details). The dashed blue lines indicate a 95% confidence interval. For country weights, see Appendix Table E.4.

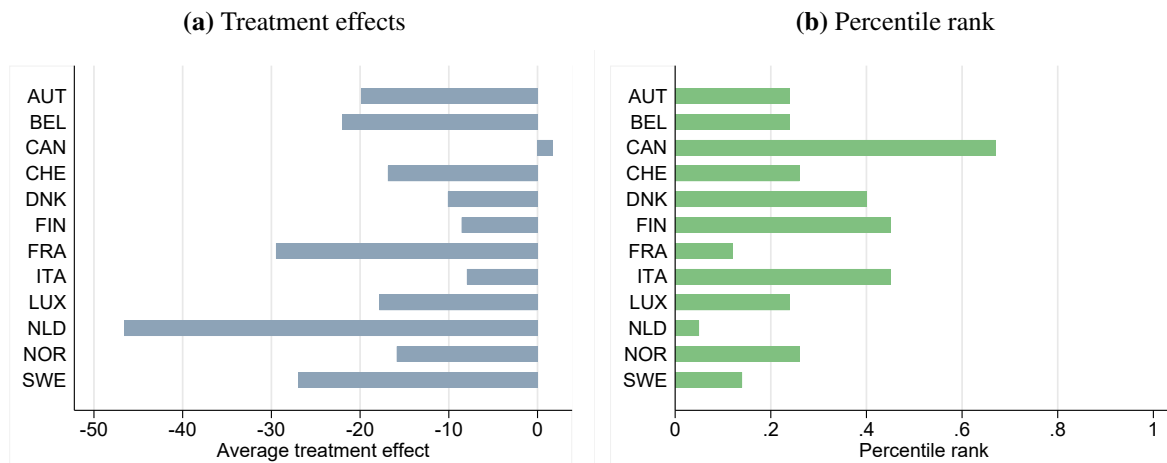
**Table IV:** Pooled treatment effect for SO<sub>2</sub>. 1980-1994.

	Average	Specific years	
	1980-1994	Year 5	Year 10
Treatment effect (mean)	-18.371	-22.603	-21.543
Treatment effect (HL)	-16.583	-20.000	-18.333
Mean rank	0.294 ***	0.236 ***	0.288 ***
95% CI (low)	-23.750	-27.083	-27.083
95% CI (high)	-6.417	-11.417	-11.417

Notes: The mean treatment effect  $\hat{\beta}_{e1}$  is estimated from equation 5. HL refers to the Hodges-Lehmann (HL) point estimate, see Section 4.3.2. Critical values for the mean percentile rank are derived from the simulation procedure described in Appendix C.2. Inverting this rank gives the 95% confidence intervals (CI). See Table E.1 for yearly treatment effects. Significance level: 1%: .302, 5%: .349, 10%: .375. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 6.1.1 Country estimates

Figure 2 shows the average treatment effects and percentile ranks for each individual country. Looking at all countries ratifying the Helsinki protocol, only one country (Canada) experienced a non-negative treatment effect (see Figure 2a). While the individual treatment effects are never statistically significant at a 5% level when using a two-sided test (see Figure 2b), the pooled treatment effects is significantly different from zero at a 1% level (see Table IV). Appendix G presents the development in emissions for each single treated country and the corresponding synthetic control.

**Figure 2:** Country-specific treatment effects. SO<sub>2</sub>. 1980-1994

Notes: Panel (a) shows the average, country-specific treatment effect derived from equation 4. Panel (b) shows the country-specific, percentile ranks derived from equation 6. See Table II for country acronyms. See Table G.1 in the Online Appendix for country-specific treatment effects in table format. See Figure G.1 the Online Appendix for figures showing the country-specific treatment effects over time and placebo runs.

### 6.1.2 Match quality and weights

The validity of the estimated treatment effect depends on how closely the synthetic control approximates key predictors for the treated countries. From Figure 1, we see that the pre-intervention development in normalized emissions is not a perfect match. However, from Figure 1a and Figure 1b we see that the 5-6 years before the treatment show a good match.

In addition to normalized emissions, I use pollution per capita, GDP per capita and fossil fuel share as predictors. Table V shows the average match of the four predictors. While the yearly development in normalized emissions show some deviations, the average over the pre-treatment period is very similar. For SO<sub>2</sub> emissions per capita and GDP per capita, the average values are somewhat higher for the treated countries.<sup>59</sup> However, the difference in pollution levels is almost entirely driven by one country: Canada. For all other countries, the pollution level per capita is about the same (see Appendix Figure E.1). For GDP per capita, the difference is driven by five countries: Canada, Switzerland, Denmark, Luxembourg, and Norway. If I restrict the treatment group to countries with a close match on GDP, the average treatment effect actually increases to -23%. The favorable treatment effect is hence not driven by the richest countries in the sample.

Appendix Table E.4 gives an overview of weights assigned to each country in the donor pool.

<sup>59</sup>Note that the difference is much smaller compared to the average of the entire donor pool (see Table III).

The countries most often used to construct synthetic controls are the United States, Iceland, New Zealand, and the United Kingdom.

**Table V:** Match of predictors. SO<sub>2</sub>. 1980.

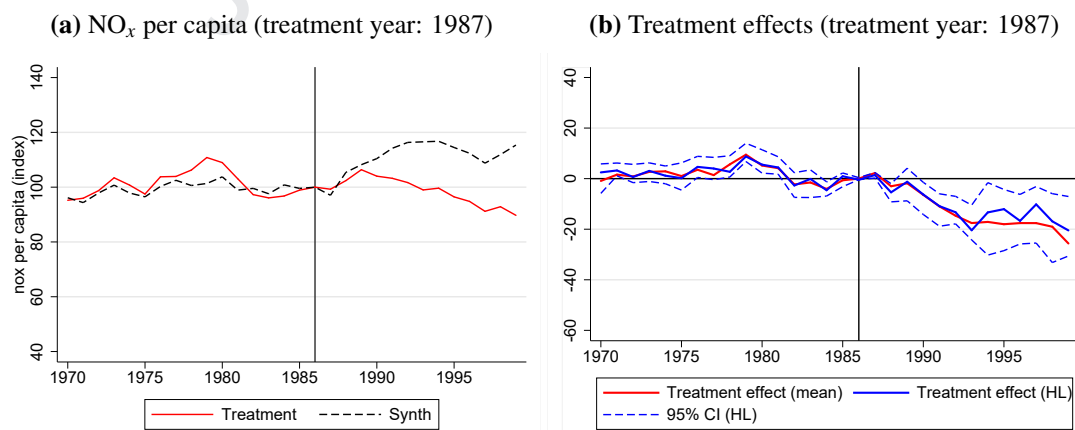
Predictor	Treated	Synthetic
SO <sub>2</sub> per capita (normalized) (1970-1980)	104.42	102.42
SO <sub>2</sub> per capita (level) (1980)	73.92	69.83
GDP per capita (1980)	27787.75	24320.25
Fossil share (1980)	81.92	76.58

*Notes:* Figure E.1 in the Online Appendix shows the match of predictors for each individual country. Table E.7 in the Online Appendix shows the weight assigned to each predictor.

## 6.2 Effects of the Sofia protocol on NO<sub>x</sub> emissions

The estimated effect of the Sofia protocol on NO<sub>x</sub> emissions is reported in Figure 3 and Table VI.<sup>60</sup> Using the baseline year as the intervention date (1987), NO<sub>x</sub> emissions in the treated countries were on average 12-13% lower in the post-treatment period compared to the synthetic control. The treatment effect is significant at a 1% level. Five years after the intervention, emissions were around 11% lower than the synthetic control, while this difference increases to 17-18% ten years after the intervention.

**Figure 3:** Effects of the Sofia protocol on NO<sub>x</sub> emissions



*Notes:* Panel (a) shows the development in emissions for the treatment group (red, solid line) and the synthetic control group (black, dashed line). Emissions in the year before treatment is normalized to 100. Panel (b) shows yearly treatment effects. The solid red line corresponds to the average, yearly treatment effects  $\hat{\alpha}_{e1t}$  estimated from equation 5. The solid blue line indicates the HL point estimate (see Section 4.3.2 for details). The dashed blue lines indicate a 95% confidence interval. For country weights, see Appendix Table E.5.

<sup>60</sup>See Appendix Table D.2 for a list of the treated and donor countries.

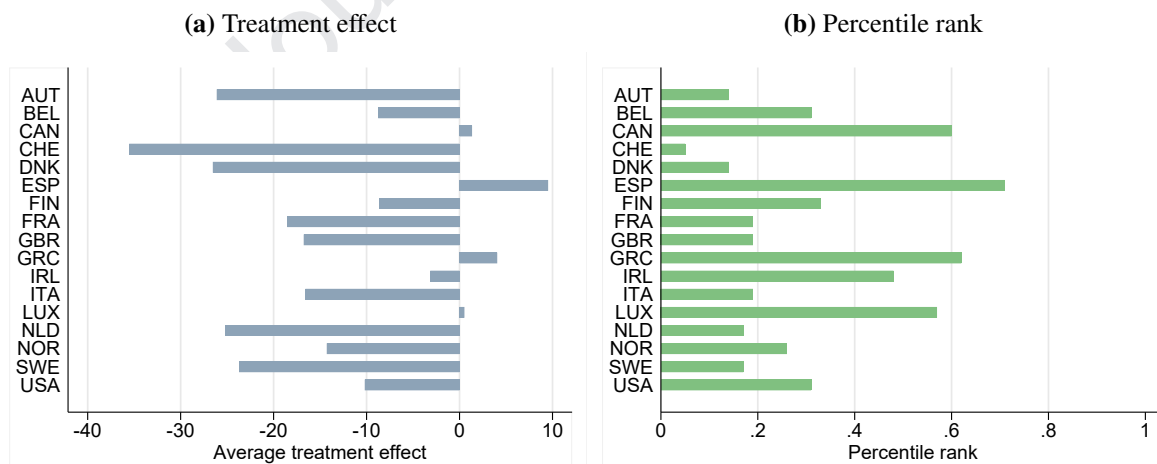
**Table VI:** Pooled treatment effect for NO<sub>x</sub>.

	Average	Specific years	
	1987-1999	Year 5	Year 10
Treatment effect (mean)	-12.858	-10.966	-17.570
Treatment effect (HL)	-11.824	-10.882	-16.706
Mean rank	0.319 ***	0.301 ***	0.314 ***
95% CI (low)	-20.000	-18.824	-18.824
95% CI (high)	-5.765	-6.000	-6.000

Notes: The mean treatment effect  $\bar{\beta}_{e1}$  is estimated from equation 5. HL refers to the Hodges-Lehmann (HL) point estimate, see Section 4.3.2. Critical values for the mean percentile rank are derived from the simulation procedure described in Appendix C.2. Inverting this rank gives the 95% confidence intervals (CI). See Table E.2 for yearly treatment effects. Significance level: 1%: .333, 5%: .375, 10%: .396. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 6.2.1 Country estimates

Figure 4 gives an overview of the country-specific treatment effects and corresponding percentile ranks. With the exception of four countries (Canada, Spain, Greece, and Luxembourg), the treatment group experienced a decrease in emissions compared to the synthetic control. Although country estimates are never statistically significant at a 5% level (using a two-sided test), the pooled treatment effect is significantly different from zero at a 1% level (see Table VI).

**Figure 4:** Country-specific treatment effects. NO<sub>x</sub>. 1987-1999

Notes: Panel (a) shows the average, country-specific treatment effect derived from equation 4. Panel (b) shows the country-specific, rank-based p-values derived from equation 6. See Table II for country acronyms. See Table G.4 in the Online Appendix for country-specific treatment effects in table format. See Figure G.2 the Online Appendix for figures showing the country-specific treatment effects over time and placebo runs.

In the years after 1987, two specific policies were introduced in Europe that could be interpreted as a result of the international cooperation facilitated by the LRTAP framework. First,



in 1988 a European Communities<sup>61</sup> Directive was introduced which required large combustion plants to significantly reduce SO<sub>2</sub> and NO<sub>x</sub> emissions compared to 1980 levels.<sup>62</sup> Second, in 1993 a EU directive targeting cars was launched, requiring all cars sold within the European Union to be fitted with a catalytic converter, which lowers NO<sub>x</sub> emissions.<sup>63</sup> Both Directives are in line with the goals stated in the Sofia Protocol, and could hence be seen as an instrument for countries to meet their emission targets. Specifically, Article 2 of the Sofia protocol states that countries need to introduce emissions standards or other pollution control measures to stationary sources (such as large combustion plants), and Article 4 mandates countries to facilitate the circulation of vehicles equipped with catalytic converters. At the same time, it could also be the case that the Sofia protocol simply reflected coordination efforts within the European Union that would have emerged also in the absence of the protocol. To test to what extent EU directives are driving the result, I compare treatment effects for countries part of the EU to non-members and countries that joined the EU at the very end of the post-intervention period.<sup>64</sup> Results show that EU countries actually experienced a smaller decline in emissions (-10%) relative to non-EU countries (-17%), suggesting that emission reductions were not contingent on the two Directives.<sup>65</sup>

### 6.2.2 Match quality and weights

The average development in NO<sub>x</sub> emissions for the synthetic control tracks the treatment group fairly well in the pre-intervention period, see Figure 3b. Although the pre-treatment difference is significantly different from zero in some years, the treatment effect fluctuates around zero for the 16 year period prior to the intervention. Table V shows the average match of the four predictors. Both normalized NO<sub>x</sub> emissions and NO<sub>x</sub> levels show a close match. GDP per capita,

<sup>61</sup>European Union (EU) from 1993.

<sup>62</sup>*The Council Directive 88/609/EEC of 24 November 1988 on the limitation of emissions of certain pollutants into the air from large combustion plants.* For more information, see <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31988L0609>.

<sup>63</sup>The policy also diffused to other non-EU countries, and in 1995 Iceland required all new vehicles to have a catalytic converter.

<sup>64</sup>The following 10 treated countries were part of the European Commission/European Union for the entire post-intervention period: Belgium, Denmark, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Spain, the UK. The following 7 countries are non-members or joined the EU in 1995: Austria, Switzerland, Finland, Norway, Sweden, Canada, USA.

<sup>65</sup>Average treatment effects can be derived from Figure 4 or Appendix Table G.4.

however, is higher for the treated countries, while the opposite is the case for the fossil fuel share. Excluding the six treated countries with the poorest match on GDP per capita leaves the average treatment estimate nearly unchanged, showing that results are not driven by the richest countries. Donor countries frequently used as controls include Japan, United Arab Emirates, and Hong Kong.<sup>66</sup>

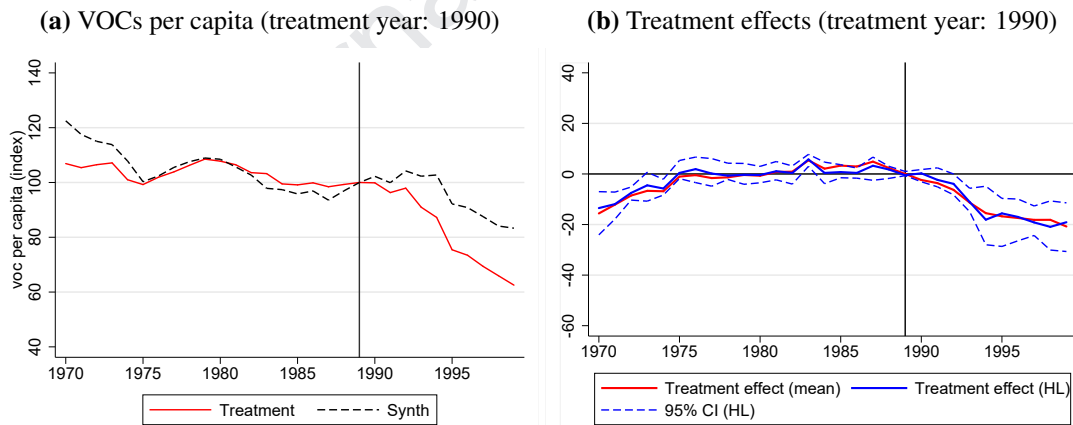
**Table VII:** Match of predictors. NO<sub>x</sub>. 1987.

Predictor	Treated	Synthetic
NOX per capita (normalized) (1970-1987)	100.82	99.18
NOX per capita (level) (1987)	43.35	43.06
GDP per capita (1987)	28929.82	24395.41
Fossil share (1987)	77.47	86.53

*Notes:* Figure E.1 in the Online Appendix shows the match of predictors for each individual country. Table E.8 in the Online Appendix shows the weight assigned to each predictor.

### 6.3 Effects of the Geneva protocol on VOCs

**Figure 5:** Effects of the Geneva protocol on emissions of VOCs



*Notes:* Panel (a) shows the development in emissions for the treatment group (red, solid line) and the synthetic control group (black, dashed line). Emissions in the year before treatment is normalized to 100. Panel (b) shows yearly treatment effects. The solid red line corresponds to the average, yearly treatment effects  $\hat{\alpha}_{e1t}$  estimated from equation 5. The solid blue line indicates the HL point estimate (see Section 4.3.2 for details). The dashed blue lines indicate a 95% confidence interval. For country weights, see Appendix Table E.6.

Figure 5 depicts the estimated yearly effects of the Geneva protocol on emission of VOC.<sup>67</sup>

Using the baseline year as the treatment date (1990), the development in emissions for the

<sup>66</sup>Note that more countries are included in the treatment group under the Sofia protocol compared to the Helsinki protocol, such as the UK and the U.S. This means that these countries are no longer in the donor pool, and we are left with “less similar” countries to choose among. For information on which countries in the donor pool are used to construct synthetic controls, see Table E.5 in the Appendix.

<sup>67</sup>See Appendix Table D.2 for a list of the treated and donor countries.

treatment group and the synthetic control diverge right after the intervention. From Figure 5b, we see that the average treatment effect increases over time. In the post-intervention period, emissions reductions for treated countries were on average 13 % larger than for the synthetic counterfactual (see Table VIII), and the effect is significant at a 1% level. Five years after the intervention emissions were 15-18% lower than the synthetic control, while this difference grows to 19-21 % ten years after the intervention.

**Table VIII:** Pooled treatment effect for VOCs.

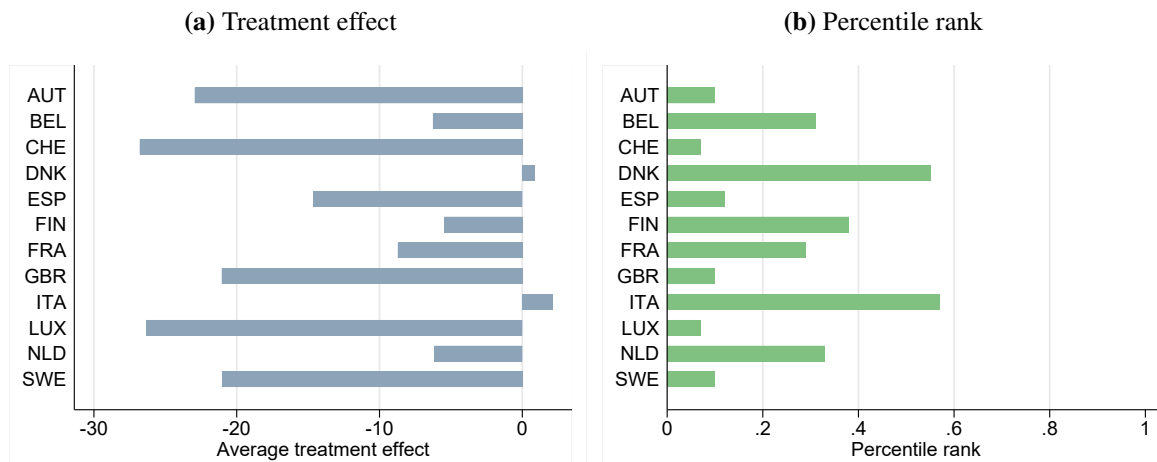
	Average	Specific years	
	1990-1999	Year 5	Year 10
Treatment effect (mean)	-13.033	-15.491	-20.779
Treatment effect (HL)	-12.917	-18.083	-19.083
Mean rank	0.248 ***	0.268 ***	0.266 ***
95% CI (low)	-21.417	-28.000	-28.000
95% CI (high)	-7.417	-4.833	-4.833

Notes: The mean treatment effect  $\bar{\beta}_{el}$  is estimated from equation 5. HL refers to the Hodges-Lehmann (HL) point estimate, see Section 4.3.2. Critical values for the mean percentile rank are derived from the simulation procedure described in Appendix C.2. Inverting this rank gives the 95% confidence intervals (CI). See Table E.3 for yearly treatment effects. Significance level: 1%: .302, 5%: .349, 10%: .375

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 6.3.1 Country estimates

**Figure 6:** Country-specific treatment effects. VOCs. 1990-1999



Notes: Panel (a) shows the average, country-specific treatment effect derived from equation 4. Panel (b) shows the country-specific, rank-based p-values derived from equation 6. See Table II for country acronyms. See Table G.7 in the Online Appendix for country-specific treatment effects in table format. See Figure G.3 the Online Appendix for figures showing the country-specific treatment effects over time and placebo runs. Norway is excluded from the treatment group due to an extreme development in VOCs.

Almost all countries experienced a decline in emissions relative to the synthetic control. Exceptions are Denmark and Italy. The percentile ranks are generally small - although none of the country-specific treatment effects are statistically significant at a 5 % level.

### 6.3.2 Match quality and weights

Looking at Figure 5, we see that the synthetic control tracks the treatment group relatively closely in the 15 years prior to intervention. The average pollution level per capita in 1990 is also very similar for the treated countries and the synthetic controls, see Table IX. Again, it is hard to find a close match on the GDP predictor as ratifying countries are systematically richer than those in the donor pool. Japan is weighted heavily in the construction of the synthetic control countries.<sup>68</sup>

**Table IX:** Match of predictors. VOCs. 1990.

Predictor	Treated	Synthetic
VOC per capita (normalized) (1970-1990)	103.00	104.75
VOC per capita (level) (1990)	51.17	48.75
GDP per capita (1990)	32521.33	25739.83
Fossil share (1990)	74.92	88.83

*Notes:* Figure E.1 in the Online Appendix shows the match of predictors for each individual country. Table E.9 in the Online Appendix shows the weight assigned to each predictor.

## 6.4 Comparing emission reductions to stated targets

How do emission reductions compare to the actual targets in each of the protocols? The Helsinki protocol committed ratifiers to reduce SO<sub>2</sub> emissions by at least 30% compared to 1980 levels, as soon as possible or by 1993. Actual emissions were 49% lower in 1993, while the constructed counterfactual shows a reduction of 27%. This implies two things: first, that ratifying countries reduced emissions well beyond the target, and second, that emissions would have declined also in the absence of the Helsinki protocol, but not nearly as much. Out of the 49% decline in SO<sub>2</sub> emissions, around 22 percentage points was induced by the Helsinki protocol. For the Sofia protocol, the stated goal was to reduce NO<sub>x</sub> emissions to 1987 levels by 1994. On average,

<sup>68</sup>For information on which countries are used to construct synthetic controls, see Appendix Table E.6.

actual emissions were at the same level in 1994 as in 1987, meaning that they were on target, while the synthetic counterfactual indicates a 20% increase in emissions in 1994 compared to 1987. The empirical findings hence suggest that the Sofia protocol curbed a rise in emissions. Under the Geneva protocol, countries could choose between different emission reduction targets, where one of them was a 30% reduction by 1999 compared to 1990. Actual emissions were 37% lower in 1999, while the synthetic control shows a 19% reduction, suggesting that the Geneva protocol contributed to an additional reduction of around 18 percentage points.

## 6.5 Robustness checks

This section reports results from a battery of robustness checks, including using alternative inference procedures, changing the predictor set and donor pool, and redefining the intervention date to later years. I also estimate treatment effects of the Helsinki protocol using the same DiD approach as in [Aakvik and Tjøtta \(2011\)](#), with the aim of disentangling the source of discrepancy between results.

### 6.5.1 Using alternative inference procedures

Table X shows the statistical significance of the main results under the four different inference procedures described in Appendix C and Section 4.1. The first procedure is the one used in the main analysis (*Independent ranks (Irwin-Hall, discrete)*). The second procedure uses continuous values instead of discrete to construct cut-off values, and results in a significance level of 5% instead of 1% in the case of the Helsinki protocol. As described in Appendix C, generating continuous values will lead to a distribution of percentile ranks that is slightly skewed to the left compared to a case with discrete values. This means that the cut-off values resulting from this procedure will be artificially low, and the robustness check hence constitutes a conservative test of the statistical significance of results. An alternative approach to conduct inference is to randomly permute the treatment status in a dataset with both actually treated countries and donor countries (*Independent ranks (randomization)*). Iterating the permutation procedure one million times gives me a distribution of percentile ranks, which are used to evaluate statistical significance. Using this time-consuming inference procedure results in the same significance

levels as in the main analysis for all three protocols.

**Table X:** Comparing significance levels under different inference procedures

	Helsinki (1980-1994)	Sofia (1987-1999)	Geneva (1990-1999)
Treatment effect (mean)	-18.371	-12.858	-13.033
Treatment effect (HL)	-16.583	-11.824	-12.917
Mean rank	0.294	0.319	0.248
<i>Significance levels under different procedures:</i>			
Independent ranks (Irwin-Hall, discrete)	1%	1%	1%
Independent ranks (Irwin-Hall, continuous)	5%	1%	1%
Independent ranks (randomization)	1%	1%	1%
Dependent ranks (randomization)	10%	1%	1%

*Notes:* Significance levels are from a two-sided test. See Sections C.1-C.4 in the Appendix for details on the inference procedures. All cut-off values are listed in Table C.3.

In a last robustness check, I relax the assumption of independent ranks (*Dependent ranks (randomization)*). Specifically, I address the potential problem that treated countries are assigned many of the same donors, and that this implies rank dependency across space. To address this issue, I repeat the permutation procedure described above, but constrain the “treatment group” to consist of countries located geographically close to each other. For the Helsinki protocol, the procedure generates a significance level of 10%, while the estimated effect of the other two protocols are still significant at a 1% level.<sup>69</sup>

### 6.5.2 Changing the predictor set

To see how the choice of predictor set and predictor weights influence treatment effects, I run five different robustness checks. Results are presented in Figure 7 and Table XI, while information on country and predictor-specific weights are provided in Online Appendix H.

In the first robustness check (R1), I use the same predictor set as in the main specification, but use a data-driven procedure to assign predictor weights.<sup>70</sup> This means that each predic-

<sup>69</sup>While spatially correlated errors might pose a threat to inference, and hence warrants a robustness check, it is worth noting that autocorrelated errors are not likely to affect inference. This is because test statistics are based on a placebo/permutation procedure and a persistent outcome variable should therefore not matter for the ranking of treatment effects. Note further that while autocorrelation might lead to biased estimates in a fixed effects regression model with lagged dependent variables as covariates (Nickell, 1981), autocorrelation is not likely to pose a threat to the unbiasedness of the synthetic control estimator (Gobillon and Magnac, 2016; O’Neill et al., 2016).

<sup>70</sup>Specifically, I use the built-in optimization algorithm provided by Abadie et al. (2011).

tor might be given a different weight in the construction of synthetic controls across countries. Compared to the main results, the optimization approach leads to a larger weight on past emission trends at the expense of the remaining predictors. The reweighing of predictors, however, has minor effects on the estimated average treatment effects, suggesting that results are not sensitive to the decision to fix predictor weights in the main specification.

In robustness checks R2-R4, I test the sensitivity of results to adding the following four predictors: population growth, GDP growth, industry's share of GDP, and forest cover.<sup>71</sup> Including additional predictors lead to relatively similar results as the main specification. The main reason for this is that the additional predictors get assigned a relatively small weight, suggesting that they do not add much information beyond what is already captured by past emissions, GDP per capita, and the fossil fuel share. Lastly, I restrict the predictor set to the two emission variables only (R5). Trimming the predictor set leads to a larger treatment effect for SO<sub>2</sub>, while the effects on NO<sub>x</sub> and VOCs are relatively similar. The larger effect for SO<sub>2</sub> is primarily caused by a change in the composition of the synthetic control towards lower income countries. This robustness check hence illustrates the importance of including GDP as a predictor to ensure that the synthetic control reflects a similar level of economic development as the treated unit.

### 6.5.3 Examining the role of spillovers and complementarities

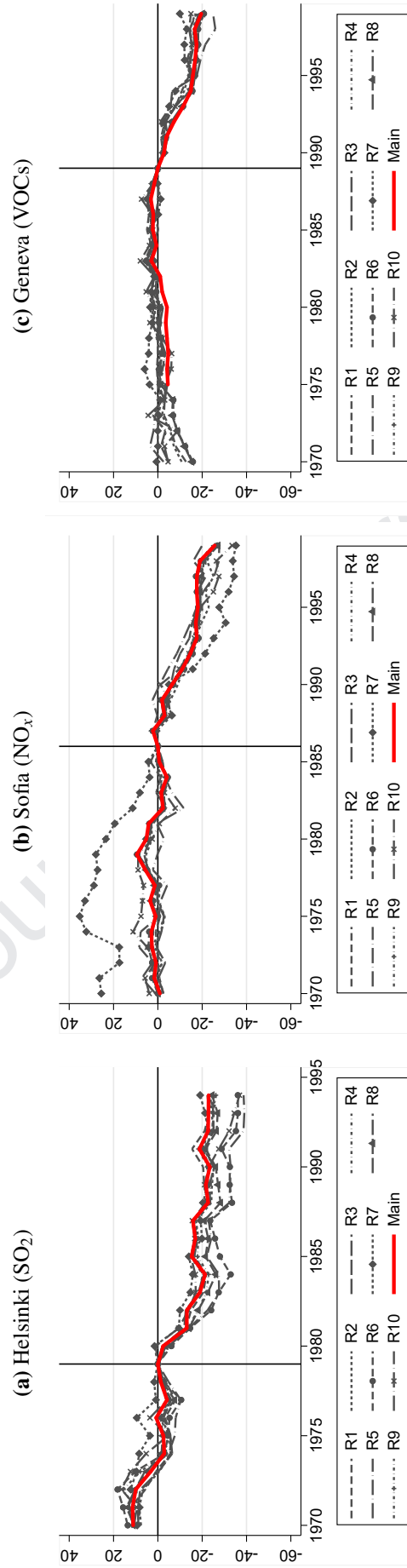
As discussed throughout the paper, estimated treatment effects might be underestimated if there are positive technology or policy spillovers to nearby control countries, or if there are complementarities between LRTAP protocols. In the main specification such positive spillovers are likely less of a problem compared to previous studies. The reason for this is that previous studies use a sample of European, LRTAP countries only, while this paper uses a sample of countries from all over the world. However, to further examine this issue, I run several robustness checks where I impose different restrictions on the donor pool.

I start by excluding all non-ratifying LRTAP countries from the set of donor countries (R6). This will likely mitigate potential problems of positive spillovers across countries within the LRTAP framework – at the expense of leaving us with less similar countries in the donor pool.

---

<sup>71</sup>The predictors are assigned weights based on the data-driven optimization algorithm.

**Figure 7: Robustness checks: changing the predictor set and donor pool**



*Notes:* Each figure shows average, yearly treatment effects  $\hat{\alpha}_{e,t}$ , estimated from equation 5 for ten different robustness checks. For a description of the robustness checks see Table XI and the text. See Appendix H for country and predictor weights used in each robustness check.



**Table XI: Robustness checks: changing the predictor set and donor pool.**

	Changing the predictor set					Changing the donor pool					
	Baseline	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
<b>Panel A: Helsinki (SO<sub>2</sub>)</b>											
Treatment effect (mean)	-18.371	-17.894	-19.197	-18.772	-18.771	-27.449	-27.799	-16.726	-22.061	-19.649	-19.979
Treatment effect (HL)	-16.583	-11.750	-12.333	-13.583	-15.000	-25.583	-19.167	.	-18.000	-17.917	-18.500
Mean rank	0.294 ***	0.315 **	0.315 **	0.278 ***	0.268 ***	0.179 ***	0.252 ***	.	0.229 ***	0.254 ***	0.282 ***
<b>Panel B: Sofia (NO<sub>2</sub>)</b>											
Treatment effect (mean)	-12.858	-13.156	-15.030	-11.732	-13.312	-11.851	-13.144	-20.791	-13.696	-12.858	-15.371
Treatment effect (HL)	-11.824	-11.353	-11.059	-8.412	-10.941	-5.588	-12.471	.	-11.471	-11.824	-13.882
Mean rank	0.319 ***	0.311 ***	0.305 ***	0.332 ***	0.304 ***	0.367 **	0.307 ***	.	0.337 **	0.319 ***	0.300 ***
<b>Panel C: Geneva (VOCs)</b>											
Treatment effect (mean)	-13.033	-10.155	-10.185	-11.396	-10.998	-12.689	-13.100	-8.388	-12.716	-13.033	-10.270
Treatment effect (HL)	-12.917	-9.250	-10.000	-11.250	-8.833	-16.167	-16.333	.	-14.667	-12.917	-10.417
Mean rank	0.248 ***	0.331 **	0.280 ***	0.312 **	0.329 **	0.230 ***	0.205 ***	.	0.202 ***	0.248 ***	0.312 **

Notes: Each panel (A-C) shows results from 10 different robustness checks (R1-R10), as well as the baseline result. Critical values for the mean percentile rank are derived from the simulation procedure described in Appendix C.2. See Appendix Table C.3 for cut-off values and Appendix H for country and predictor weights.

R1: Baseline sample. Baseline predictor set. Data driven approach to assign weights to predictors.

R2: Baseline sample. Add GDP growth and population growth as predictors. Data driven approach to assign weights to predictors.

R3: Baseline sample. Add GDP growth, population growth, and forest cover as predictors. Data driven approach to assign weights to predictors.

R4: Baseline sample. Add GDP growth, population growth, forest cover, and industry share of GDP as predictors. Data driven approach to assign weights to predictors.

R5: Baseline sample. Only keep pre-intervention values of the outcome variable (pollution) as predictors.

R6: Baseline predictor set. Drop non-ratifying LRTAP countries from donor pool.

R7: Baseline predictor set. Only keep non-ratifying LRTAP countries in the donor pool.

R8: Baseline predictor set. Drop countries in the bottom 40% of the income distribution

R9: Baseline predictor set. Keep countries of all income levels, as well as countries experiencing highly volatile emissions

R10: Baseline predictor set. Drop (literately) the top two donor countries (SO<sub>2</sub>: United States and New Zealand, NO<sub>x</sub>: Japan and New Zealand, VOC: Japan and Ireland).

Significance levels for SO<sub>2</sub>: 1%: .302, 5%: .349, 10%: .375

Significance levels for NO<sub>x</sub>: 1%: .333, 5%: .375, 10%: .396

Significance levels for VOCs: 1%: .302, 5%: .349, 10%: .375

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Results from the estimation show that the absolute treatment effect of the Helsinki protocol increases from -18% to -28%, while the increase in the HL estimate is more modest. Countries assigned a large weight in the construction of synthetic controls include Australia, New Zealand, Japan, and Chile (see Appendix H). While the larger treatment estimate might suggest that positive spillovers downward-bias treatment effects in the main specification, it is challenging to credibly quantify the size of such a bias. Instead, the robustness check gives an indication of the sign of the bias. Treatment effects of the Sofia and Geneva protocols are less affected, which might not be that surprising as non-ratifying LRTAP countries make up a small part of the donor sample in the main specification.

In R7, I do the opposite of R6 and restrict the donor pool to non-ratifying LRTAP countries only. This specification is more similar to the sample definition used in previous literature.<sup>72</sup> If there are positive spillovers between protocols or other favorable effects of participating in the broader LRTAP framework, such effects would be differenced out in this specification. The resulting estimate could hence be interpreted as a lower bound effect of the international cooperation facilitated by the LRTAP framework. Using the restricted and relatively small donor pool of countries, I find that the absolute treatment effect of the Helsinki protocol is reduced from -18.3% to -16.7%. For the Geneva protocol, the estimated treatment effect is now around 5 percentage points lower. Note that the estimated effects of the Sofia and Geneva protocols rely almost entirely on 1-2 donor countries (Cyprus and Ireland), which means that any post-intervention shock or spillovers to these control countries would severely impact the treatment estimates. The pre-treatment match for the Sofia protocol is also poor, meaning that the estimated (larger) treatment effect is not reliable (see Figure 7). Due to the small number of countries in the donor pool, I cannot perform inference on these estimates. However, the robustness check suggests that failing to account for positive spillovers is likely to lead to an underestimation of the treatment effect.

---

<sup>72</sup>However, I also exclude countries heavily affected by the fall of the Soviet Union, which are included in many of the previous studies.

#### 6.5.4 Examining the role of different income levels

Another potential problem is that treated countries have on average a higher GDP per capita compared to donor countries. In the main estimation, I aim to mitigate this concern by including GDP per capita as a predictor, and by dropping the poorest quintile from the donor pool. I have also shown that treatment effects are not driven by the richest countries in the treatment group. To further examine the issue of different income levels, I drop the *two* poorest quintiles from the donor pool (R8). This increases the effect of the Helsinki protocol slightly (from -18.3% to -22%), showing that the main result is not driven by the use of relatively poor countries in the donor pool. The effects of the Sofia and Geneva protocol are very similar to the baseline. The small changes to the estimated treatment effects are not surprising as low-income countries are rarely given a positive weight in the baseline estimation.<sup>73</sup> In R9, I instead relax the restrictions on the donor pool and include countries of all income levels as well as countries experiencing highly volatile emissions. This has little effect on the treatment estimates, but lowers the statistical significance level of the Helsinki protocol from 1% to 5%. Overall, the results from R8 and R9 suggest that including GDP as a predictor is sufficient to ensure we compare countries of relatively similar income levels, and that trimming the donor pool by income does not add much to improve comparability between treated countries and the synthetic controls.

#### 6.5.5 Excluding top donor countries

Lastly, we might worry that the synthetic controls rely on a couple of key donor countries, which could make the treatment estimates vulnerable to specific shocks to these countries. For instance, New Zealand is often among the top donor countries for SO<sub>2</sub>, while Japan is often chosen as a donor when estimation effects on NO<sub>x</sub> and VOCs. In R10, I exclude (iteratively) the top two donor countries used to construct synthetic controls in each of the three protocols.<sup>74</sup> While excluding countries that received a large weight in the main specification implies sacrificing some goodness of fit, it also allows us to evaluate to what extent treatment effects are driven by a few key donor countries (Abadie et al., 2015). Excluding the top two donors slightly

<sup>73</sup>In Appendix J I also estimate treatment effects using emissions as a share of GDP as the outcome variable instead of emissions per capita. Normalizing emissions by GDP corroborates the main findings, although treatment effects are slightly lower.

<sup>74</sup>SO<sub>2</sub>: The United States and New Zealand, NO<sub>x</sub>: Japan and New Zealand, VOCs: Japan and Ireland.

increases estimated effects of the Helsinki and Sofia protocols, and slightly lowers the effect of the Geneva protocol while maintaining a good pre-treatment fit. The stability of the treatment effect is reassuring, as it suggests that the constructed counterfactuals are not just an artifact of a specific country combination.

#### 6.5.6 Anticipation effects or (forward-looking) self-selection?

In the main specification, I have chosen to backdate the intervention date to account for potential anticipation effects. This decision is based on the general recommendations in [Abadie \(2012\)](#), as well as the fact that negotiations typically start several years prior to the formal protocol meetings. In the case of SO<sub>2</sub>, talks about emission reductions started already at the 1979 Convention meeting, and the specific emission reduction target used in the Helsinki protocol was proposed already in 1983 ([Sliggers et al., 2004](#)). Moreover, the LRTAP organized several technology seminars, the first one in 1981, that facilitated exchange of information about abatement opportunities ([Sliggers et al., 2004](#)). These activities prior to the protocol meeting likely sent a signal to both Governments and firms about the likelihood of a future international agreement on SO<sub>2</sub>. Such anticipation effects are also discussed in [Dekker et al. \(2012\)](#), where the authors find suggestive evidence of a run-up effect of the Helsinki protocol on patenting behavior.

While backdating of the treatment date is warranted only if countries do indeed reduce emissions as a response to the anticipated policy, there is also the possibility that what we observe are not anticipation effects but rather a self-selection of countries into the protocol based on declining emissions. If this is the case, it would be better to shift the intervention date closer to the formal decision by countries to sign and ratify the protocol. Another potential problem is that countries might self-select into protocols based on the *expected* future emission development. This type of forward-looking selection hypothesis implies that even if we have chosen the “correct” intervention date, and we have managed to construct a synthetic control that closely matches the pre-intervention development of the treated county, the treatment estimates might still suffer from (forward-looking) self-selection bias. Ruling out such a hypothesis is challenging, as it requires knowledge of countries’ expectations about future emission developments.

In an attempt to address these two concerns, I have done the following. First, to address

**Table XII:** Changing the intervention year

	Treatment effect		
	Mean	HL	Mean rank
<i>Panel A: Helsinki protocol (SO<sub>2</sub>)</i>			
1980: Baseline year of the protocol	-18.371	-16.583	0.294 ***
1981: Intermediate year	-18.803	-12.333	0.288 ***
1983: Intermediate year	-14.397	-14.917	0.319 **
1985: Helsinki meeting	-11.684	-10.500	0.339 **
1986: Median ratification year	-11.422	-11.000	0.296 ***
1987: Helsinki protocol enters into force	-6.993	-4.417	0.365 *
<i>Panel B: Sofia protocol (NO<sub>x</sub>)</i>			
1987: Baseline year of the protocol	-12.858	-11.824	0.319 ***
1988: Sofia meeting	-16.969	-15.706	0.333 **
1990: Median ratification year	-13.029	-9.353	0.374 **
1991: Sofia protocol enters into force	-9.666	-4.588	0.378 *
<i>Panel C: Geneva protocol (VOC)</i>			
1990: Baseline year of the protocol	-13.033	-12.917	0.248 ***
1991: Geneva meeting	-19.893	-19.250	0.234 ***
1994: Median ratification year	-10.338	-9.833	0.204 ***
1997: Geneva protocol enters into force	-5.670	-6.667	0.290 ***

*Notes:* Table shows estimated treatment effects from 14 different estimations. Each row represents results from a separate estimation. The year in the first column indicates the year chosen as the intervention year. Critical values for the mean percentile rank are derived from the simulation procedure described in Appendix C.2. For more details on results and weights, see Appendix I.

Significance levels for SO<sub>2</sub>: 1%: .302, 5%: .349, 10%: .375

Significance levels for NO<sub>x</sub>: 1%: .333, 5%: .375, 10%: .396

Significance levels for VOCs: 1%: .302, 5%: .349, 10%: .375

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

the potential problem of choosing a too early intervention date, I examine the sensitivity of results to using three alternative definitions of the intervention year: i) the year of the protocol meeting, ii) the median ratification year, and iii) the year the protocol entered into force. For the Helsinki protocol, I also show results for two intermediate years (1981 and 1983) given the relatively long period between the baseline year and the Helsinki protocol meeting. Results from the estimations are presented in Table XII. Generally, we see that treatment effects tend to decline as we define a later intervention year. For the Helsinki protocol, the treatment effect drops from a 18-19% reduction when defining 1980 or 1981 as the intervention year, to a 11-12% reduction when opting for the Helsinki protocol meeting or the median ratification year as the intervention date. The treatment estimate further diminishes to 7% when defining entry into

force as the treatment year.<sup>75</sup>

For the Sofia and Geneva protocols, we have seen from the main results that treatment effects start to materialize a few years into the post-treatment period. Shifting the intervention date one year forward to the protocol meetings hence increase the treatment effects, while shifting it further to the median ratification year and entry into force lower effects.<sup>76</sup> While it is hard to conclude on the “correct” intervention year, all specifications yield negative treatment effects that are statistically significant at a 5 or 10% level. Still, the results illustrate the importance of carefully considering the choice of treatment year.

The potential biased caused by forward-looking self-selection is harder to address. However, one approach to partly mitigate this concern is to include all LRTAP countries in the treatment group - irrespective of whether they signed and ratified the protocol in question. This approach avoids the problem of self-selection among LRTAP countries, but obviously does not deal with forward-looking self-selection into the initial 1979 LRTAP Convention. The ‘intention-to-treat’ effects are reported in Appendix K and show treatment effects that are around 3-4 percentage points smaller in magnitude compared to the main specification. These effects could be seen as conservative estimates of the protocol-specific treatment effects.

### 6.5.7 Using a difference-in-differences approach

Lastly, I try to identify the main reason why previous studies have found no effect of the Helsinki protocol. In particular, I compare my findings to the most similar study; [Aakvik and Tjøtta \(2011\)](#). Combining a difference-in-difference (DiD) approach with data on SO<sub>2</sub> dating back to 1960, the authors find no significant effect of the Helsinki protocol on emissions. In Appendix F, I try to disentangle the main source of discrepancy by examining the role of sample, treatment year, and non-parallel trends. First, estimating a DiD model with leads and lags using the same sample and intervention year as in the main specification, I find large and negative treatment effects, but also a clear violation of the parallel trends assumption. Using 1985 as the treatment year instead of 1980 aggravates the gap in pre-treatment trends, and also lowers the estimated

<sup>75</sup>As shown in Appendix I, it is hard to construct a synthetic control that is able to closely mimic the sharp drop in emissions in the treatment group in the years after 1981.

<sup>76</sup>Note that using entry into force as the intervention year for the Geneva protocol leaves us with a post-intervention period of only 3 years.

treatment effect. Restricting the control group to LRTAP countries only has minor effects on the estimate. Lastly, I follow [Aakvik and Tjøtta \(2011\)](#) and include country-specific linear time trends. Doing so completely wipes out the treatment effect, suggesting that the way non-parallel trends are dealt with is the main reason for their conclusion of no effect. While we clearly need to address the non-parallel trends, including linear time trends may not be a suitable approach as it seems to absorb most of the treatment effect. Using a synthetic control method offers an alternative approach to the problem of different pre-treatment trends.

## 7 Concluding remarks

Understanding the potential of international agreements to mitigate cross-border environmental externalities is crucial to guide policy makers towards instruments that make a real difference. At the same time, evaluating impacts of multilateral agreements is methodologically challenging, and to date there are few empirical studies that credibly establish causal relationships between ratification status and subsequent environmental outcomes, such as air pollution.

In this study, I revisit three large-scale pollution protocols on  $\text{SO}_2$ ,  $\text{NO}_x$  and VOCs implemented in the 1980s and early 1990s, with the aim of establishing causal impacts of the protocols on emissions. By combining a newly assembled dataset with a method for constructing synthetic counterfactuals, I am able to address several potential problems associated with previous empirical examinations, such as non-parallel emission trends, spillovers, and anticipation effects. Results from the empirical estimation suggest that the international protocols induced sizable emissions reductions of all three pollutants. For ratifying countries,  $\text{SO}_2$  emissions were on average 22% lower than the synthetic control group ten years after the intervention, while the corresponding numbers for  $\text{NO}_x$  and VOC emissions were 18% and 21%, respectively. Additionally, I find that the protocol targets went beyond a “business as usual” scenario (i.e., the synthetic control), and that countries on average either met or exceeded the protocol targets.

These results suggest that international protocols can be an effective tool to induce countries to lower their emissions, contrasting the often gloomy predictions from the game theoretical literature. The findings also illustrate how different pre-intervention trends and anticipation effects can bias estimated treatment effects, if these are not accounted for in a suitable manner.



Specifically, the comparative analysis in this paper suggests that previous studies on the Helsinki protocol have tended to underestimate the favorable effects on SO<sub>2</sub> emissions.

Although I address several shortcomings of previous studies, there are still other, more inherent features of international protocols that make it challenging to recover causal estimates. In particular, accounting for all types of direct and indirect spillovers in a highly complex and globalized economy, is close to impossible. If ratifying countries lower their emissions by e.g., reducing the use of high sulfur coal, this could potentially lead to increased emissions in non-ratifying countries via trade flows. At the same time, higher technology adoption by ratifying countries might stimulate technological development and diffusion, potentially inducing emissions reductions also in non-ratifying countries. In order to fully account for all positive and negative spillovers, one would need detailed, global data on technology adoption and product-level trade flows between countries. Such an analysis is beyond the scope of this paper, but might be a fruitful avenue for future research.

## References

- Aakvik, A. and Tjøtta, S. (2011). Do collective actions clear common air? The effect of international environmental protocols on sulphur emissions. *European Journal of Political Economy*, 27(2):343–351.
- Abadie, A. (2012). Using synthetic controls to evaluate an international strategic positioning program in uruguay: Feasibility, data requirements, and methodological aspects. Harvard University, mimeo.
- Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.
- Abadie, A., Diamond, A., and Hainmueller, J. (2011). SYNTH: Stata module to implement Synthetic Control Methods for Comparative Case Studies. Statistical Software Components, Boston College Department of Economics.
- Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.
- Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1):113–132.



- Almer, C. and Winkler, R. (2017). Analyzing the effectiveness of international environmental policies: The case of the Kyoto protocol. *Journal of Environmental Economics and Management*, 82:125 – 151.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics : an empiricist's companion*. Princeton University Press.
- Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, pages 878–894.
- Benchenkroun, H. and Long, N. V. (2012). Collaborative environmental management: A review of the literature. *International Game Theory Review*, 14(04):1240002.
- Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001.
- Bratberg, E., Tjøtta, S., and Øines, T. (2005). Do voluntary international environmental agreements work? *Journal of Environmental Economics and Management*, 50(3):583–597.
- Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, 95(5):1549–1561.
- Dekker, T., Vollebergh, H. R. J., de Vries, F. P., and Withagen, C. A. (2012). Inciting protocols. *Journal of Environmental Economics and Management*, 64(1):45–67.
- Dinda, S. (2004). Environmental kuznets curve hypothesis: A survey. *Ecological Economics*, 49(4):431 – 455.
- Dube, A. and Zipperer, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies. *IZA DP No. 8944*.
- Finus, M. and Tjøtta, S. (2003). The Oslo protocol on sulfur reduction: the great leap forward? *Journal of Public Economics*, 87(9-10):2031–2048.
- Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.
- Hodges Jr, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, pages 598–611.
- Hoel, M. (1992). International environment conventions: the case of uniform reductions of emissions. *Environmental and Resource Economics*, 2(2):141–159.
- Houghton, K. A. and Naughton, H. T. (2016). *Comparative Law and Economics*, chapter Chapter 18: International environmental agreement effectiveness: A review of empirical studies. Elgar.
- Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., and Schaaf, E. (2011). Emissions database for global atmospheric research, version v4.2 (time-series). European Commission, Joint Research Centre (JRC).
- Jørgensen, S., Martín-Herrán, G., and Zaccour, G. (2010). Dynamic games in the economics and management of pollution. *Environmental Modeling & Assessment*, 15(6):433–467.

- JRC (2012). The emissions database for global atmospheric research (EDGAR). European Commission, Joint Research Centre (JRC)/PBL Netherlands Environmental Assessment Agency. Available at: [edgar.jrc.ec.europa.eu/datasets\\_list.php?v=42](http://edgar.jrc.ec.europa.eu/datasets_list.php?v=42). Version 4.2.
- Mideksa, T. K. (2013). The economic impact of natural resources. *Journal of Environmental Economics and Management*, 65(2):277 – 289.
- Mitchell, R. B. (2002-2015). International environmental agreements database project (version 2014.3). Available at: <http://iea.uoregon.edu/>. Date accessed: 20 April 2015.
- Murdoch, J. C., Sandler, T., and Sargent, K. (1997). A tale of two collectives: sulphur versus nitrogen oxides emission reduction in Europe. *Economica*, 64(254):281–301.
- Murdoch, J. C., Sandler, T., and Vijverberg, W. P. M. (2003). The participation decision versus the level of participation in an environmental treaty: a spatial probit analysis. *Journal of Public Economics*, 87(2):337–362.
- Naughton, H. T. (2010). Globalization and emissions in Europe. *The European Journal of Comparative Economics*, Vol. 7(No. 2):pp. 503–519.
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49 (6):1417–26.
- O'Neill, S., Kreif, N., Grieve, R., Sutton, M., and Sekhon, J. S. (2016). Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Services and Outcomes Research Methodology*, 16(1-2):1–21.
- Pinotti, P. (2015). The economic costs of organised crime: Evidence from southern Italy. *The Economic Journal*, 125(586):203–232.
- Powell, D. (2017). Synthetic control estimation beyond case studies: Does the minimum wage reduce employment?
- Ringquist, E. J. and Kostadinova, T. (2005). Assessing the effectiveness of international environmental agreements: The case of the 1985 Helsinki protocol. *American Journal of Political Science*, 49(1):86–102.
- Sliggers, J., Kakebeeke, W., and Nations, U. (2004). Clearing the air: 25 years of the convention on long-range transboundary air pollution.
- Stern, D. I. (2006). Reversal of the trend in global anthropogenic sulfur emissions. *Global Environmental Change*, 16(2):207 – 220.
- The World Bank (2015). The World Bank Indicators. Available at: <http://data.worldbank.org/indicator/>. Data accessed: July 2015.
- UNECE (2015). Protocols. Retrieved from: [www.unece.org/env/lrtap/status/lrtap\\_s.html](http://www.unece.org/env/lrtap/status/lrtap_s.html). Accessed: August 2015.
- Vollenweider, J. (2013). The effectiveness of international environmental agreements. *International Environmental Agreements: Politics, Law and Economics*, 13(3):343–367.