Journal of Environmental Economics and Management 95 (2019) 227–256

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

The Kyoto protocol: Empirical evidence of a hidden success ☆

Nada Maamoun *

Institute for Law and Economics, University of Hamburg, Gemany

ARTICLE INFO

Article history:
Received 16 January 2018
Revised 1 April 2019
Accepted 4 April 2019
Available online 10 April 2019

JEL classification:
C2
F53
H87
Q53
Q54

Keywords:
Treatment effects
Kyoto protocol
International public goods
Air pollution
Global warming
Generalized synthetic controls

ABSTRACT

How effective was the Kyoto protocol? International Environmental Agreements (IEA) have been on the rise over the past four decades; however, thus far their effectiveness is controversial. In view of the conflicting results found in the related literature, this paper addresses its effectiveness by utilizing for the first time the generalized synthetic control method (GSCM) to compare the emissions of the industrialized countries with a "No- Kyoto" counterfactual scenario that represents the expected emissions in the absence of the protocol. This method facilitates a robust comparison between treated and control countries as done by Almer and Winkler (2017) and account for the multiple treated units as done by Grunewald and Martinez-Zarzoso (2016), so as to capture the collective nature of the protocol. Results show that the protocol was successful in reducing the emissions of the ratifying countries approximately by 7% below the emissions expected under a "No-Kyoto" scenario, confirming the importance of accounting for the collective nature of the agreement.

1. Introduction

Climate change and the problem of global warming have climbed to the forefront of global concern and become the center of scientific as well as global public debate. Not only due to its increasingly negative effects on current ecological and economic systems, but also due to the uncertainty that surrounds the causes as well as future effects of global warming (Mitchell, 2003; Libecap, 2014). The air pollution levels have been exponentially increasing and current emissions of anthropogenic greenhouse gases are 70% more than they were 40 years ago (Barrett et al., 2014).

Taking into account the trans-boundary nature of air pollution, global collective action is needed to target such a global problem (Feeny et al., 1990; Arce M. and Sandler, 2001; Aakvik and Tjøtta, 2011; Wiener, 2007; Livermore and Revesz, 2017). Therefore, it comes as no surprise that ever since the late 1970's the number of international agreements signed and policies set trying to reduce air pollution has been exponentially increasing. One of the most prominent treaties is the Kyoto protocol

☆ I would like to thank Yiquing Xu for providing the R-codes, the simulation data and his guiding comments. I would also like to thank Lamis Sherif, Adriana Espes, Max Mantei, Jose Reis, Julia Samwer, Oliver Englisch, Marwa el Abhar, Ana Odorovic, Mohamed Fakhreddin, Stephan Michel, Jerg Gutmann, Konstantinos Pilpilidis, Claudia Schwirplies, Michael Berlemann, Stefan Voigt and Franziska Weber for their helpful comments and advice. I also gratefully acknowledge the DFG (German research foundation) funds for the graduate school of "The Economics of Internationalization of Law".

* Johnsallee, 35, 20148, Hamburg, Germany.

E-mail address: nada.maamoun@ile-graduateschool.de.

https://doi.org/10.1016/j.jeem.2019.04.001
0095-0696/

(referred to as KP or the protocol thenceforth), which is a part of the United Nations' Framework Convention on Climate Change (UNFCCC) (Barrett, 1994; Mickwitz, 2003; Mitchell, 2003; Faure and Lefevere, 2012; Vollenweider, 2013).

The KP has legally binding targets that require the industrialized countries to reduce their emissions to be on average 5.2% less than their emission levels in the base year 1990 during the first commitment period (United Nations, 1998; Breidenich et al., 1997; Aldy and Stavins, 2013). One of the important features of the protocol is its international approach. With more than 190 countries, the KP was truly seeking collective action. To facilitate cooperation among the countries, the protocol offered mechanisms such as the International Emission trading (IET), the Joint Implementation (JI) and Clean Development Mechanisms (CDM) through which the countries can achieve their targets jointly or separately (Böhringer, 2003; Aldy and Stavins, 2008; Faure and Lefevere, 2012). This international approach is very important when it comes to the analysis of its effectiveness.

Previous literature has analyzed the effectiveness of the KP in reducing harmful emissions using different empirical methods and has provided conflicting results. On the one hand, (Aichele and Felbermayr, 2013; Grunewald and Martinez-Zarzoso, 2016) show evidence that the ratification of the protocol had an effect on the reduction of the emissions. On the other hand (Almer and Winkler, 2017) show that the protocol had non significant emission reduction in the parties included. While these contributions have their merits empirically, the methods are not without shortcomings when it comes to the analysis of the protocol considering the nature of air pollution and the protocol.

Given the specificities of air pollution -that air is transboundary- as well as the nature of international agreements and the rules and obligations associated with it, the current empirical evidence does not sufficiently address the effectiveness of the protocol. The analysis of the emissions levels of those countries has to take into consideration the selection bias that arises form the voluntary nature of the protocol as well as the collective aspect of the agreement. Hence the need for adequate evaluation of the protocol has become essential, especially with the increased number of agreements being signed and ratified (Barrett, 1994; Mickwitz, 2003; Vollenweider, 2013; Livermore and Revesz, 2017).

The purpose of this paper is to add to the existing empirical literature on the effectiveness of the KP, by using a newly introduced method - the generalized synthetic control method (GSCM)- introduced by Xu (2017). The GSCM facilitates the analysis of the protocol's effectiveness and accounts for the selection bias as well as the collective nature of international agreements, while at the same time overcoming the shortcomings of the methods previously used. In essence, the use of GSCM facilitates a robust comparison between treated and control countries as done by Almer and Winkler (2017) and account for the multiple treated units as done by Grunewald and Martinez-Zarzoso (2016), so as to capture the collective nature of the protocol.

Analyzing data on the GHG emissions using the GSCM, results show that the KP has been successful in reducing the emissions of the countries that were legally bound by the protocol by approximately at least 6% below their expected business-as-usual (BAU) levels. The results are more in line with (Aichele and Felbermayr, 2013; Grunewald and Martinez-Zarzoso, 2016) rather than the results shown by Almer and Winkler (2017). To ensure robustness, several checks were run to check whether the criticism against the protocol was in place or not, and the results show that overall the KP has consistently led to a significant reduction in the GHG emissions of the legally bound parties.

The contribution of this paper is twofolds. It introduces the usage of the GSCM for the first time to analyze the effectiveness of international agreements; in general such analysis has been subject to problems such as endogeneity and Xu (2017) provides a method -GSCM- that facilitates such an analysis. In addition it contributes to the existing literature on the effectiveness of the KP as a collective action by proposing that the protocol's effectiveness is more fairly assessed when dealing with participating countries as a group rather than as individual units, given the transboundary nature of air pollution as well as the collaborative nature of the protocol. Additionally, the papers shed some light on some of the specifications of the Kyoto protocol that were subject to criticism such as "Hot-air"[2] and leakage and whether such criticism was in place or not.

The paper is divided into five sections. The following section gives a brief background on the KP as well as previously used methods of analysis. The third section discusses the empirical strategy and the specifications of the GSCM. The fourth section analyses the results and the last section concludes.

2. Background

2.1. The Kyoto protocol

The Kyoto protocol is part of UNFCCC targeting the reduction of harmful gases and it is one of the most infamous IEAs. The KP has been criticized before it came into force (Barrett, 1998; Böhringer and Vogt, 2004) and even more after its first commitment period ended in 2012 and the problem of global warming was still escalating (Aichele and Felbermayr, 2012; Nordhaus, 2015). The protocol was criticized for several reasons, the first and main criticism was the differentiation among countries in terms of legal obligation under the UNFCCC's principle of "Common but differentiated responsibilities".

The protocol was ratified by 192 countries and in this sense it is an international agreement, however, it sets legally binding targets to only a group of countries known as Annex B countries (industrialized countries). This differentiation is assumed to

[2] Assigned amounts of the former soviet union countries, which were deemed to be more than their expected BAU.

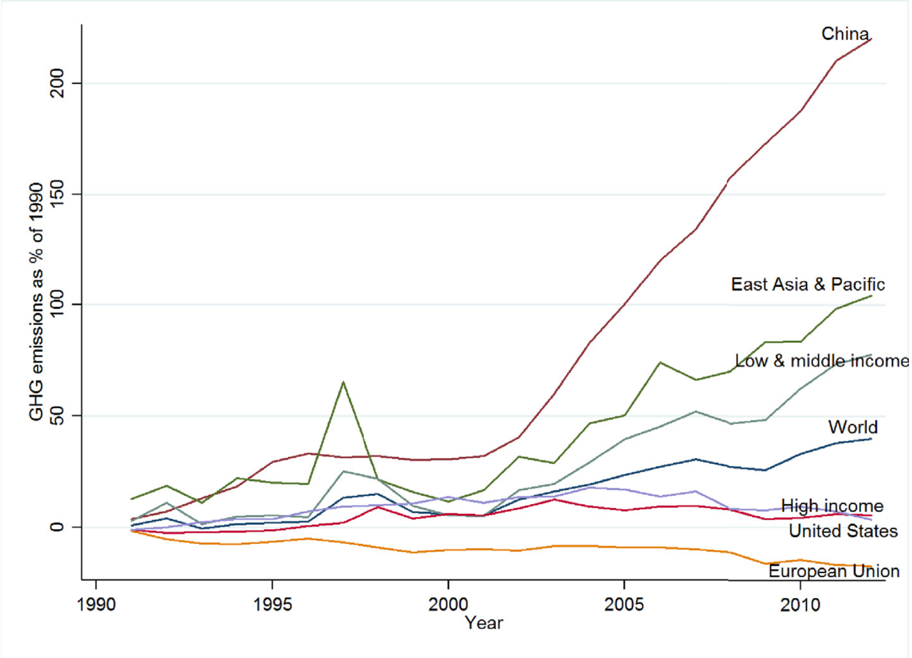

Fig. 1. The GHG emissions as % change of 1990 emission levels.

have undermined the effectiveness of the protocol as it means that the agreement is in fact a sub-global one (Wiener, 2007). The legally binding targets stated in the protocol require the industrialized countries to reduce their emissions to be on average 5.2% less than their emission levels in the base year 1990[3] during the first commitment period (United Nations, 1998; Breidenich et al., 1997; Aldy and Stavins, 2013). The differentiation in the obligations of the parties -while sensible at the time- limits the legal obligation within the borders of a small number of countries and leads to the problem of leakage (Böhringer and Vogt, 2004; Nordhaus, 2015). The leakage problem lies in the fact that air is trans-boundary and shifting the emissions across borders does not lead to any effect; in fact it could be counterproductive if the developing economies exceed the production of the industrialized world (Aldy and Stavins, 2013).

Another criticism of the protocol is the amounts assigned to the former soviet-union countries (FSU). These amounts were larger than their expected BAU emissions levels -known as “hot-air”- which opened the opportunity of increasing emissions or selling the extra assigned amounts, without any reduction in emissions levels. According to the protocol, in the cases of non-compliance the non-complying party will face a deduction in the next commitment period equivalent to 1.3 times the amount of the violation. It may also be prevented from participating in the market-based mechanisms provided by the protocol. However, for this penalty to be applied it needs the consent of the noncomplying party, and that’s one of the reasons why these measures are weak (Böhringer, 2003). Furthermore, these measures do not present themselves as sufficient incentives to encourage compliance unlike some other protocols -such as the Montreal protocol[4] - that uses trade sanctions on violating parties (Sunstein, 2007).

Additionally, the first commitment period of the protocol was set to be 5 years 2008–2012, which given the nature of the climate problem is not enough (Aldy and Stavins, 2008), especially since carbon dioxide (CO_2) and other GHG tend to remain in the atmosphere for decades. While this period may encourage participation, it was a major disincentive when it comes to investment decisions of technological nature such as investment in cleaner technology or renewable energy facilities (Barrett, 1998). While the criticism of the KP, may deem the protocol weak or inefficient, the protocol’s effectiveness should not be disregarded or undermined. It might be the case that the protocol has been successful in preventing a worse scenario from taking place.

Current evidence of global emission shows that there is an increasing trend in emission levels. This seems to be in line with the expectations of the protocol’s opponents. As shown in Fig. 1 the global GHG emissions are increasing relative to 1990 levels and they are driven mainly by the emissions of the developing world and in specific the Asian countries. On the other hand, the emissions of the EU exhibit a negative trend and the high-income countries emissions are stable.

[3] The base year is set to be 1990 for most industrialized countries except for the newly capitalized countries (eastern European countries), who were allowed to choose a different year.

[4] Montreal protocol aimed at reducing ozone depleting substances and is considered to be one of the effective international agreements targeting air pollution.

This provides evidence that the criticism of the protocol on the grounds of it being not sufficiently inclusive was valid and that leakage might have been a prominent problem. However, a simple comparison between the emissions of the industrialized countries legally bound by the protocol (European Union (EU) and some high income countries) and the rest of the world's emissions would not be accurate. To examine the protocol's effectiveness more fairly it would be better to compare between the current GHG emissions of the industrialized countries with the expected BAU emissions that would have occurred in the absence of the protocol. Looking at it this way might show that the KP was actually successful at least in preventing a worse-off situation, where the industrialized countries might have had higher emission levels.

2.2. Previously used methods

The general empirical approach to assess the effectiveness of an environmental agreement is to compare current pollution levels with the "would-have been" pollution rate if the policies were not imposed. The goal is to find the suitable counterfactual scenario and accordingly find the "would-have been" pollution rate in an untreated country in comparison to a country that had a policy set with the aim of pollution reduction. (Helm and Sprinz, 2000) use game theoretic models to find the counterfactual and while it is theoretically elegant it has some empirical concerns with regards to the accuracy of its predictions as well as its ability to identify causal effects over time, given that the data is subject to possible variability and many confounding variables that are not always observable as well as the presence of important spatial and temporal dimensions (Vollenweider, 2013). Turning a blind eye on these characteristics would lead to inaccuracy as well as contradiction in interpretation (Scott, 2007).

A more preferred and commonly applied method is to investigate the impact of environmental policies using regressions. However, the voluntary nature of the agreements gives rise to the presence of unobservable confounders that lead to a bias (selection bias) due to the fact that the reasons why each government voluntarily takes the step to participate in an agreement differ from one government to another and are most likely endogenous to country-specific characteristics.[5] This selection bias needs to be taken into account when evaluating the effectiveness of policies to avoid biased results, (Bennear and Coglianese, 2005; Vollenweider, 2013; Aichele and Felbermayr, 2012). To overcome the previously mentioned limitations more recent research has moved to the usage of the difference-in differences (DID) method (Vollenweider, 2013; Grunewald and Martinez-Zarzoso, 2016).

Vollenweider (2013) used the DID to test the effectiveness of the Gothenburg protocol, one of the protocols of the Long-run Transboundary Air Pollution (LRTAP) agreements.[6] The problem with the DID though is that it sets forth the assumption that the unobservable differences are constant over time, which in the case of environmental impacts is not a very accurate assumption due to the existence of confounders that vary over time. To overcome the lack of parallel trends, matching has been used before the DID estimation as in Grunewald and Martinez-Zarzoso (2016), so that they are able to make the control and treated group more similar to each other and hence limit the effects of the unobservable time varying confounders. In their paper, they tested the effect of ratifying the KP on CO_2 emissions of the industrialized countries that were legally bound by the protocol to reduce their emissions. Their results show a significant 7% reduction in CO_2 emissions in the countries bound by the protocol. They use the propensity scores of the ratification of the KP to match countries with commitments to the closest country that has no commitments. By doing that they reduce the unobservable differences between the treated and control group, hence reducing the bias that would result from the un-parallel trends between the two groups. They also show how the mean values of the predictors used in the matched control group are very close to that of the treated group. But while the use of matching has its merits and indeed in their paper Grunewald and Martinez-Zarzoso (2016) show that their matched group has emissions trends parallel to the treated group, that is not always the case and the use of matching does not necessarily ensure that the trends before treatment will be parallel and accordingly the estimates would not be fully valid (Bennear and Coglianese, 2005; Grunewald and Martinez-Zarzoso, 2016; Xu, 2017).

In this case the use of the synthetic control method (SCM) seems to be a more empirically feasible method for application due to the fact that it does not assume the presence of parallel trends (Abadie et al., 2010). It formulates from the control group (non-treated group of countries) a synthetic counterfactual that behaves very similar to the original treated unit in the years prior to the treatment, thus facilitating accurate comparison between the treated unit and its hypothetical behavior in the case of no treatment (Vollenweider, 2013). Almer and Winkler (2017) analyzed the effectiveness of the KP on 15 countries (out of 36 countries mentioned in Annex B in the protocol) using the SCM, where they found negative but non significant effects on the treated countries. The main limitation of this paper lies in the fact they do not account for the fact that the KP is a collective action initiative that provided some market-based tools such as the IET scheme as well as others to facilitate the reduction of the participating group. Analyzing countries individually would disregard these facilitating tools and hence provide inaccurate results. This could also be one of the reasons that contributed to the vast difference between their results and those presented by Grunewald and Martinez-Zarzoso (2016).

[5] Controlling for some factors may reduce the bias but is still insufficient as the confounders are usually not known and hence regression models would not provide the most accurate results.

[6] That targets the reduction of 4 harmful polluting gases, namely SO_2, NO_x, NH_3 and VOC.The paper shows that the Gothenburg's protocol had negative but insignificant effects on the emissions of SO_2 and NO_x emissions.

Given that the purpose of this paper is to analyze the effectiveness of the KP as international agreement aiming at collectively reduce trans-boundary air pollution, estimating the effects on individual countries would not serve this aim. There have been suggestions by Abadie et al. (2010) to aggregate the treated units into a single treated unit and aggregate the control units into groups. However one problem with this approach is that statistical significance is difficult to detect given the reduced number of control units after aggregation. Kreif et al. (2016) propose an extension of weighing the control units to match the average pretreatment outcomes of treated unit. Nevertheless, Xu (2017) provides a more accurate and efficient method, the GSCM, that allows for having multiple treated units. It also provides frequentist uncertainty estimates that are more reliable than the placebo tests used in SCM, which makes the GSCM more efficient as well as easily interpretable compared to the original SCM. So in view of the conflicting results presented by Grunewald and Martinez-Zarzoso (2016) and Almer and Winkler (2017), this paper addresses KP's effectiveness by utilizing the GSCM that would both facilitate a robust comparison between treated and control countries as done by Almer and Winkler (2017) and account for the multiple treated units as done by Grunewald and Martinez-Zarzoso (2016), so as to capture the collective nature of KP.

3. Empirical strategy

3.1. Data

The dataset consists of 153 countries, 34 treated countries and 119 control countries. Out of the 36 parties in Annex B that originally ratified the protocol, only 34 countries are used due to the limited data availability of the remaining 2 countries: Monaco and Liechtenstein. The data covers 18 years from the period 1995 till 2012.[7] The treatment period is 2005 for all of the 34 parties except Australia as it ratified the protocol in 2007.[8]

The treatment period was set to be 2005 (enforcement period) rather than 2008 (commitment period) for several reasons. First, by 2005 most of the industrialized countries had ratified the protocol and given that the enforcement period is closer to the period of ratification, which is a process that involves the voters and politicians as well as domestic policy adjustments and hence essentially binds the policy makers to the protocol, it is more feasible to choose 2005 as the treatment period (Aichele and Felbermayr, 2012; Grunewald and Martinez-Zarzoso, 2016). This is also in line with previous empirical studies on Montreal Protocol as well as the Helsinki protocol (part of LRTAP), which show that parties started adopting policies that aim at reducing their emissions before the enforcement period commenced (Murdoch and Sandler, 1997; Ringquist and Kostadinova, 2005). Second, long-lived gases like the GHG remain in the atmosphere for a while, CO_2 can remain in the atmosphere for decades. Hence the earlier a country reduces its emissions, the easier it will be for them in later years to maintain lower emission levels, especially given the that the KP reduction targets are expected to be reached by the end of the 1st period- 2012- and not annually (Barrett, 1998). Third, using the enforcement period will facilitate the comparison between the results estimated by the GSCM and those found in previous literature.[9]

The output variable used is GHG excluding land use and forestry (LUCF) (in natural logarithm) from CAIT Climate Data Explorer 2015 provided by World Resources Institute (WRI)[10]. The CAIT provides data on the historical emissions of 6 major GHGs for 185 countries and the EU for the period 1990–2012.[11] To ensure completeness and accuracy of the datasets, the WRI complements the data provided by the UNFCCC by compiling data from research center, government agencies and non-governmental sources (CAIT Climate Data Explorer, 2015). The reason for using GHG without the inclusion of LUCF is to have a clearer picture of the protocol's effect on the GHG reduction without the inclusion of sinks, as it has been one of the criticisms of the protocol.

The control variables are used so as to account for the different emitting behavior across countries and were chosen following previous literature that addresses the effectiveness of environmental agreements (Bennear and Coglianese, 2005; Morley, 2012; Vollenweider, 2013; Grunewald and Martinez-Zarzoso, 2016; Almer and Winkler, 2017).[12] Three control variables are used. First, the real GDP per capita is used to account for the productivity of a country. Second, based on the Grossman and Krueger (1995) findings, the squared real GDP per capita is used to capture the non-linear relationship between the GHG air pollution and the income per capita. Third, the population size is used as proxy for the emitting behavior of the country in terms of

[7] In previous empirical literature that addresses KP, a longer period is used. In the Grunewald and Martinez-Zarzoso (2016) paper, the period analyzed is 1992–2009 for 32 treated countries and 138 control countries. In Almer and Winkler (2017) the period analyzed is 1980–2011, using 15 treated countries from Annex B and a control group of High and upper middle income countries. Given the lack of data available for some countries, especially the FSU, and the lack of data on GHG prior to 1990, the period under analysis in this paper starts from the year 1995 rather than earlier. However, in the appendix replication of both papers are presented, and the results do not differ much.

[8] The list of the countries can be found in Appendix A.

[9] Namely, Grunewald and Martinez-Zarzoso (2016) and Almer and Winkler (2017).

[10] Data available on: http://cait.wri.org/historical.

[11] For the countries that had change in their borders over the past years, the CAIT uses a consistent methodology to estimate previous emission levels based on the emissions in the 5 years after the formation (in the case of separation such as ex-soviet union countries) of those countries. As for the unification cases (such as Germany) the methodology is more simple, where the emissions are simply added to each other.

[12] Mainly based on the paper by Grossman and Krueger (1995) 'Economic Growth and the environment'; where the growth of income was shown to have a significantly negative effect on air pollution levels as long as this income is below $8000 (1985 dollars) if the income is more than that, their results do not show a significant negative relationship between the income growth and air quality.

consumption (Grunewald and Martinez-Zarzoso, 2016). The control variables are from the World Bank's World Development Indicators (WDI) transformed into their natural logarithm form.

3.2. Generalized synthetic control method

3.2.1. Why the GSCM?

The GSCM is used for causal inference using time-series cross-sectional data. It is used in this paper to estimate the effect of the KP on the emissions of industrialized countries by formulating a counterfactual from the control group. This counterfactual is formulated in a way such that it behaves very similarly (in terms of emissions) to the industrialized countries so that a comparison of performance post treatment could accurately reflect the effect of the treatment. In this sense the GSCM is similar to the idea of the SCM by Abadie et al. (2010) to formulate a counterfactual unit by using information from the non-treated group of units (control group). What the GSCM adds is that it unifies the SCM with interactive fixed effects model. It computes the synthetic counterfactual semi-parametrically. It formulates the counterfactual based on a linear interactive fixed effects (IFE) model that interacts unit specific intercept with time varying coefficients in 3 steps.

It has several attributes that make it preferable to other methods such as SCM and the DID, which have been used in analyzing the KP in previous literature. The GSCM adds to the SCM the possibility of having several treated units and different treatment periods, which is one of the main reasons why it is preferred over the SCM in this paper; given that the treated group consists of the 36 industrialized countries mentioned in Annex B in the protocol. Second, it improves the efficiency of the SCM and enhances its interpretability as it provides uncertainty estimates such as standard errors and confidence intervals, which makes the results easier to understand. Furthermore, it conducts dimension reductions before the reweighting scheme so that the vectors reweighted are smoother over time. Third, the GSCM estimator has a cross validation procedure built-in that -given sufficient data- automatically selects the correct number of factors of the IFE model reducing the risk of over-fitting (Xu, 2017).

As for the DID, the GSCM surpasses it in the case where the assumption of the parallel trends between the treated and control units in absence of treatment doesn't hold, like in the case of the presence of unobserved time varying confounders, which is the case here. There are some variables that can affect the emissions levels of a country such as the growth levels, population size, the dependence on the industry or agriculture sector as well as other variables that are not observed that may affect the treated unit and not the control unit and vice versa. Hence, in this case we cannot assume that the industrialized (treated) countries and the control group follow same trends.

The GSCM addresses the lack of parallel trends between control and treated units, as it models the unobservable time-varying coefficients semi-parametrically by using the interactive fixed effects (IFE) model proposed in Bai (2009). In addition to that it corrects the bias of the IFE model when the treatment effects are heterogeneous across units, which is sometimes the case with the GHG emissions trends across industrialized countries.

In general the GSCM has 3 main limitations. The first limitation is that the estimates may be biased if the pre-treatment period or the number of control units are small, to be specific a pretreatment period less than 10 years or number of control units less than 40 units. This does not pose as a problem as the control group used here is 119 countries and the pre-treatment period is not less than 10 years.

A second limitation is that it cannot handle complex data generating process (DGP); either in terms of presence of structural breaks, or the presence of a dynamic relationship between the treatment and the outcome variable (such as the case of a lagged dependent variable) or different treatment intensities and multiple treatment periods. The data used is annual GHG emissions, the treatment period is either 2005 or 2007 and the treatment intensity should not be very different (at least theoretically). Accordingly the DGP should not be complex.

A third limitation is that the model specification play an important part here unlike the SCM, where the procedure would not run if those specifications are not met, the GSCM would generate results, however, they might be biased if the treated and control units do not share the same factor loadings and would lead to excessive extrapolation. This problem-if it exists can be detected when looking at the plot of the factor loadings of the treated and control group, where the plot of treated units would deviate from the plot of the control units.[13]

3.2.2. Framework

Following the frameworks previously used in causal inference, in the GSCM Xu (2017) uses the same notations to denote the outcome of a certain unit i at time t by Y_{it}. Based on the first assumption of the GSCM, that the treated and control units have to be affected by the same set of factors with no structural breaks in the observed time period, the outcome of unit i at time t is expressed as

$$Y_{it} = \delta_{it} D_{it} + X'_{it}\beta + \lambda'_i f_t + \epsilon_{it} \tag{1}$$

where D_{it} takes the value of 1 in the case that unit i at time t was legally bound by the KP. Accordingly δ_{it} is the effect the protocol on unit i at time t. X is a (k × 1) vector of observed covariates and β is a (k × 1) vector of unknown parameters. f is

[13] The plot of the loadings of the treated group should lie in the convex hull of the control group.

an (r × 1) vector of unobserved common factors representing the time varying coefficients and λ_i is an (r × 1) vector of factor loadings representing country specific intercepts. ϵ_{it} is the error term and has a mean of zero.

Based on the aforementioned functional form, the treatment effect would then be represented by the difference between $Y_{it}(1)$ and $Y_{it}(0)$, in which $Y_{it}(1)$ represents the outcome of treated unit i at time t where $D_{it} = 1$ and $Y_{it}(0)$ is the outcome of control unit i at time t. This treatment effect is captured by

$$\delta_{it} = Y_{it}(1) - Y_{it}(0), \qquad t > T_0 \tag{2}$$

The counterfactual is formulated from the pool of control units in this case the group of countries that did not have legally binding targets in the protocol, so the outcome of a unit from the control group at time t can be expressed in the following equation

$$Y_{it} = X_{it}'\beta + \lambda_i' f_t + \epsilon_{it} \tag{3}$$

and hence combining all controls together the outcome of the counterfactual is

$$Y_{co} = X_{co}'\beta + F\Lambda_{co}' + \epsilon_{co} \tag{4}$$

where Y_{co} is a (T × N_{co}) matrix, X_{co} is a three dimensional matrix (T × N_{co} × p) and Λ_{co} is (N_{co} × r) and $X_{co}'\beta$,$F\Lambda_{co}'$ and ϵ_{co} are (T × N_{co}) matrices. β, F and Δ_{co} are constrained by 2 constraints to be able to identify them; first that all factors are normalized and second they are orthogonal to each other.[14] The main aim of the GSCM is to find the effect of the treatment by finding the average difference between the treated unit(s) and its counterfactual. This difference can be expressed by the following equation

$$ATT_{(t,t>T_0)} = \frac{1}{N_{tr}} \sum_{i\in\tau} [Y_{it}(1) - Y_{it}(0)] = \frac{1}{N_{tr}} \sum_{i\in\tau} \delta_{it} \tag{5}$$

to be able to compute the treatment effect in equation (2) and then 5, the counterfactual unit is estimated by the GSCM in 3 steps.

3.2.3. Formulation of counterfactual

The general idea is to choose the model that leads to the most accurate predictions, so as to be able to find this model some of the data is withheld (a small part) and the remaining part of the data is used to make predictions for the withheld part and the model that provides the most accurate predictions is the chosen model. The 1st step is to obtain a fixed number of latent factors by estimating an IFE model using control group data

$$\begin{aligned} (\widehat{\beta}, \widehat{F}, \widehat{\Lambda_{co}}) &= \underset{\widehat{\beta},\widehat{F},\widehat{\Lambda_{co}}}{\operatorname{argmin}} \sum_{i\in\tau} (\widetilde{Y_i} - \widetilde{X_i}\beta - \widetilde{F}\widetilde{\lambda_i})'(\widetilde{Y_i} - \widetilde{X_i}\beta - \widetilde{F}\widetilde{\lambda_i}) \\ s.t. \frac{\widetilde{F'}\widetilde{F}}{T} &= I_r \ and \ (\widetilde{\Lambda_{co}'}\widetilde{\Lambda_{co}}) \end{aligned} \tag{6}$$

The second step is to estimate the factor loadings $\widehat{\lambda_i}$ (unit specific intercepts) by projecting pretreatment outcomes of the treated units(s) on the space spanned by the factors estimated in the 1st step

$$\begin{aligned} \widehat{\lambda_i} &= \underset{\widehat{\Lambda_i}}{\operatorname{argmin}} (Y_i^0 - X_i^0\widehat{\beta} - \widehat{F}^0\widetilde{\lambda_i})'(Y_i^0 - X_i^0\widehat{\beta} - \widehat{F}^0\widetilde{\lambda_i}) \\ &= (\widehat{F}^{0\prime}\widehat{F}^0)^{-1}\widehat{F}^{0\prime}(\widehat{Y}^0 - \widehat{X_i^0}\widehat{\beta}), \quad i \in \tau \end{aligned} \tag{7}$$

The 3rd step is to estimate counterfactual based on factor and factor loadings computed in step 1 and 2

$$\widehat{Y}_{it}(0) = x_{it}'\widehat{\beta} + \widehat{\lambda}_i'\widehat{f}_t, \qquad i \in \tau, t > T_0 \tag{8}$$

4. Results

4.1. Main results

The results show that the KP has been successful in reducing the GHG emissions of the industrialized countries, even with the exclusion of sinks, when compared to the counterfactual representing BAU scenario. Results shown in Table 1 show a 6.8% reduction of emissions in the treated group as compared to the counterfactual that represents the would-have-been scenario in the absence of the protocol.

[14] N_{co} is the number of control units and N_{tr} is the number of treated units.

Table 1
Effect of Kyoto protocol on GHG emissions (excluding LUCF) for industrialized countries.

Variable	*ATT/Beta coefficient*
Ln GHG (excluding LUCF)	−0.0684** (0.02)
Ln Real GDP per capita	−5047.9 (1.51e+04)
Ln (real GDP per capita) squared	2524.15 (7.59e+03)
Ln population	0.730 (0.151)
MSPE	0.00159
Unobserved factors	1
Treated units	34
Control units	119

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *LUCF* is land-use change and forestry. *MSPE* is mean squared predictor error. Years of analysis 1995 till 2012. Treatment year is 2005, except Australia it is 2007. The treated countries are found in the 1st column in Table A.1 and the control countries in A.2.

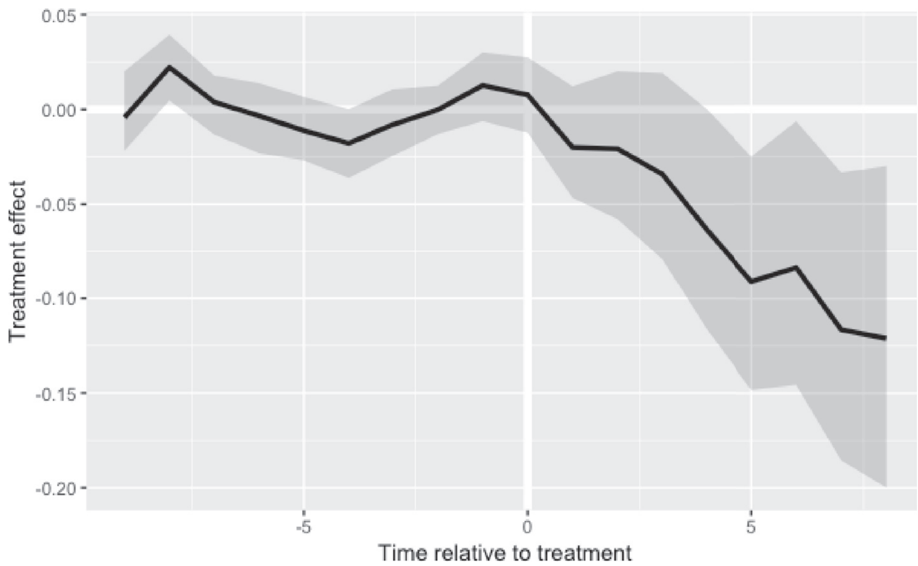

Fig. 2. The gap in the GHG emissions between the treated group and the counterfactual. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

In Fig. 2, the effect of the protocol on the emissions of the treated group is demonstrated. The counterfactual emissions are represented by the horizontal line and the treated group emissions is the deviation from the horizontal line, shown by the downward sloping curve post-treatment (the zero vertical line). The difference in the emission trends between the counterfactual and the treated group is shown in Fig. 3, where the counterfactual representing the expected BAU levels (the dashed line)- exhibits an increase in the GHG emissions compared to the actual emissions of the treated group (represented by the solid line). The cross validation scheme found 1 unobservable factor[15] to be important, so the average treatment effect on the treated countries (ATT) based on the estimations of the GSCM after conditioning on the additive fixed effects and the unobservable factors is a reduction of approximately 6.8%.

[15] Based on Xu (2017) the unobservable factors found are not directly interpretable. In this case (it is 1 factor) it may simply refer to a time trend that has a heterogeneous effect on the different unit.

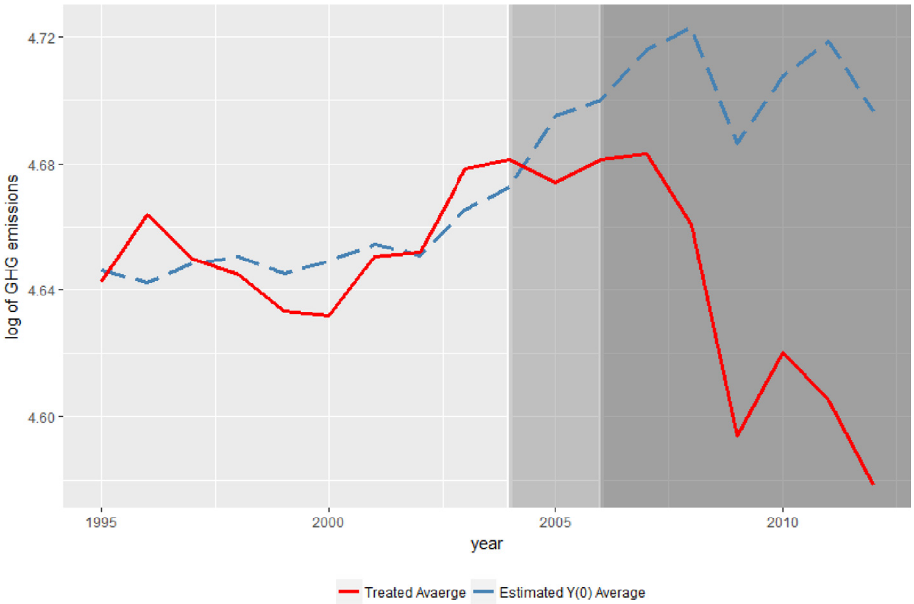

Fig. 3. The GHG emissions of the treated group and the counterfactual. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies.

In other words, the results provide evidence that the protocol has managed to succeed in reducing the emissions of the industrialized countries that ratified the protocol. The estimated reduction is approximately 7% less than their expected BAU emissions that would have occurred in the case that the protocol had not provided strict obligations and targets to be achieved for the industrialized countries to reduce their emissions. The results are surprisingly more in line with the results shown by Grunewald and Martinez-Zarzoso (2016) rather than those shown by Almer and Winkler (2017), notwithstanding the similarity between the GSCM used here and the SCM used in latter paper. This similarity goes further to show the importance of accounting for the collective nature of the protocol.

To clarify, the results should not be mistakenly interpreted that the GHG of the industrialized countries have shrunk by 6.8% than their pretreatment emissions, rather- as represented in Figs. 2 and 3- it represents a reduction compared to the emissions that were expected to take place given the protocol had not set target emissions for these countries.

4.2. Robustness checks

4.2.1. The European Union

The European community represents a large group of the industrialized countries and as a union they have other agreements such as the LRTAP and its protocols as well as directives such as the “Air quality framework” directive in 1997, the “New air quality” directive that govern and regulate air pollution and harmful gas emissions. As a consequence, the incentives presented to the EU to reduce their emissions and shift to cleaner technology is higher than other countries, even if the gases focused on in these agreements are not necessarily the same. So in order to separate the Kyoto effect from the EU effect, the European community was removed from the treated group and a new counterfactual is formed for the industrialized non-EU countries.

The results -presented in Figs. 4 and 5 as well as in the EU column in Table 2 - show that when the EU countries were removed from the treated group, the protocol still has a negative effect on the industrialized countries emissions, though the effect is significant only on a 10% level. The estimated reduction is 7% reduction approximately as compared to approximately 6.8% reduction that was estimated when the European community was included in the treated group. The results show that the removal of the EU didn’t reduce the magnitude of the effect of the protocol’s effectiveness (in fact it has slightly increased). Nonetheless, it reduced the potential significance of the result, so the results should be interpreted carefully.

Careful interpretation is essential here, especially as the results may seem counter intuitive at first glance, in the sense that the results may point to the ineffectiveness of the regional regulation in the EU. However, that is not necessarily true. First the results are less robust than the main result,[16] which makes such a direct conclusion unsupported. Second within the EU there

[16] Given that the results for this robustness check are not significant on a 5% level which is the generally an acceptable threshold for a sufficiently significant result. It is also used by the other two main comparable studies Grunewald and Martinez-Zarzoso (2016) and Almer and Winkler (2017), which makes comparison to those papers plausible.

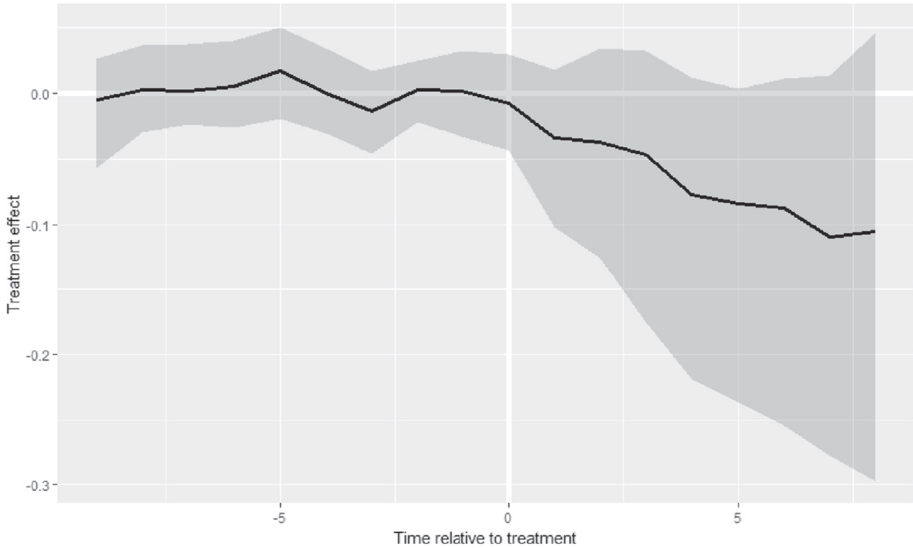

Fig. 4. The gap in the GHG emissions between the treated group and the counterfactual- EU countries excluded from treated group. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

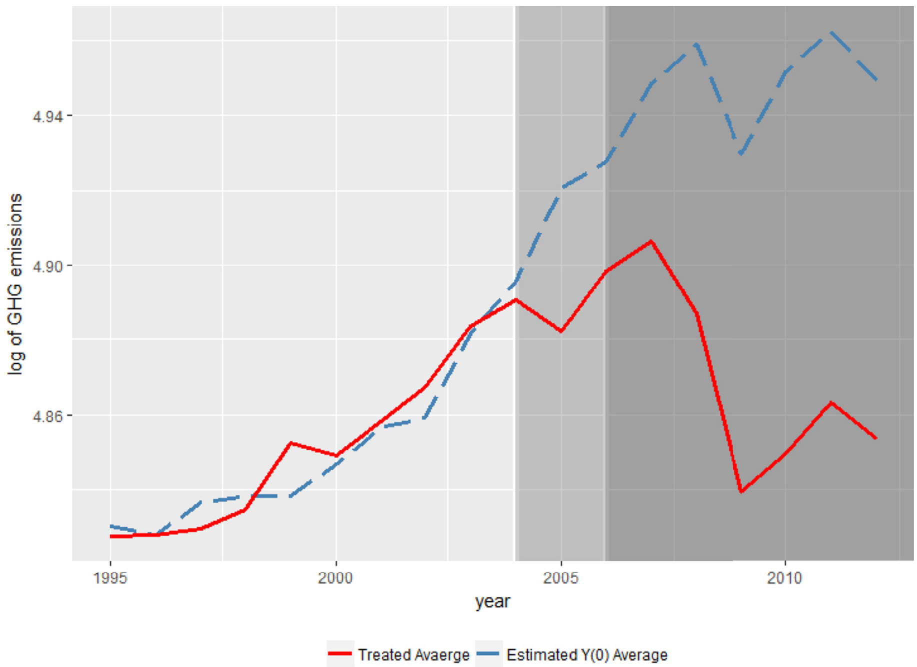

Fig. 5. The GHG emissions of the treated group and the counterfactual- EU countries excluded from treated group. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies.

are several FSU countries that were assigned 'Hot-Air' and accordingly had the right to increase emissions without violating the protocol's obligations, which may explain the less than expected reduction in the EU. The issue of 'Hot Air' is explored in details in the next section.

4.2.2. Hot air

One of the main criticisms of the KP is what is known as 'Hot Air', which refers to the assigned amounts of emissions of FSU countries that exceeded their expected BAU emissions levels. The main concern with the so-called 'Hot air' is that it may bias the results by deflating the reduction attributed to the protocol, as the FSU countries didn't have an obligation

Table 2
Robustness checks on the effect of Kyoto protocol on GHG emissions.

	EU	Hot air	CDM
Variable/ATT			
Ln GHG (excluding LUCF)	−0.071* (0.050)	−0.172** (0.075)	−0.085*** (0.020)
Ln Real GDP per capita	−5047.92 (1.52e+04)	1.0059e +04 (1.25e+04)	−1.198e +04 (1.58e+04)
Ln (real GDP per capita) squared	2524.15 (7.60e+03)	−5.029559e +03 (6.29e+03)	5.994e+03 (7.93e+03)
Ln population	0.730*** (0.166)	0.0700 (0.312)	0.055 (0.119)
MSPE	0.00100	0.00174	0.00168
Unobserved factors	1	2	1
Treated units	10	22	34
Control units	119	119	114

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *LUCF* is land-use change and forestry. *MSPE* is mean squared predictor error. *EU* refers to the robustness check in section 4.2.1. *Hot air* refers to the robustness check in section 4.2.2. *CDM* refers to the robustness check in section 4.2.3. Years of analysis 1995 till 2012. Treatment year is 2005, except Australia it is 2007. The treated countries are found in 2nd, 3rd and 1st columns in Appendix A.1, respectively and the control countries in A.2.

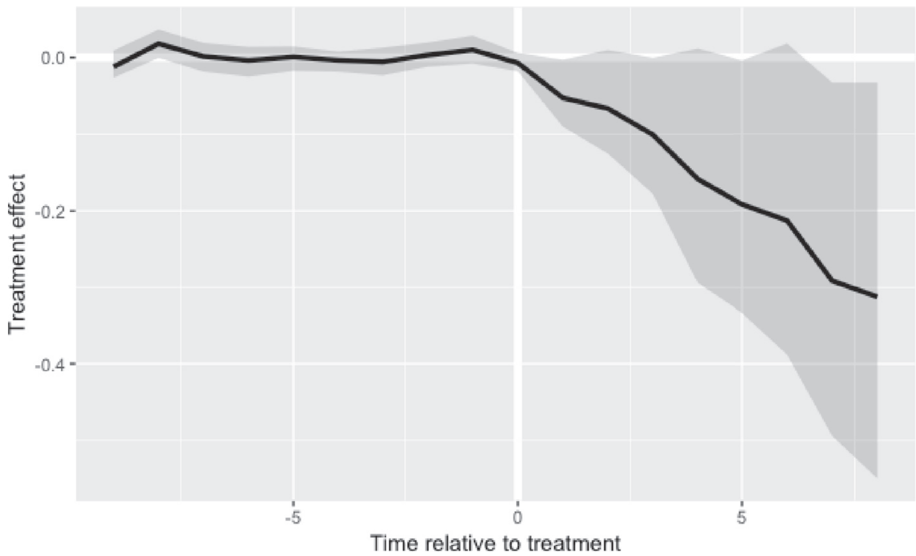

Fig. 6. The gap in the GHG emissions between the treated group and the counterfactual- FSU countries excluded from treated group. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

to reduce their emissions. Hence computing the average treatment effect while including them in the treated group would show a reduced treatment effect for the overall group as compared to the effect without their inclusion. In order to overcome this possible bias, the FSU countries were removed from the treated group so as to analyze the "net" effects of the protocol.

Results show that removing the FSU countries from the treated group increased the magnitude of reduction to be 17% (Table 2, Figs. 6 and 7) rather than the approximate 7% shown in previous results. This goes in line with the intuition that the extra assignments of emission units have motivated an increase in these countries' emissions or at the very least did not motivate any reduction of emission levels. It may also explain why the EU reductions as a group were undermined, as some of the FSU countries joined the EU.

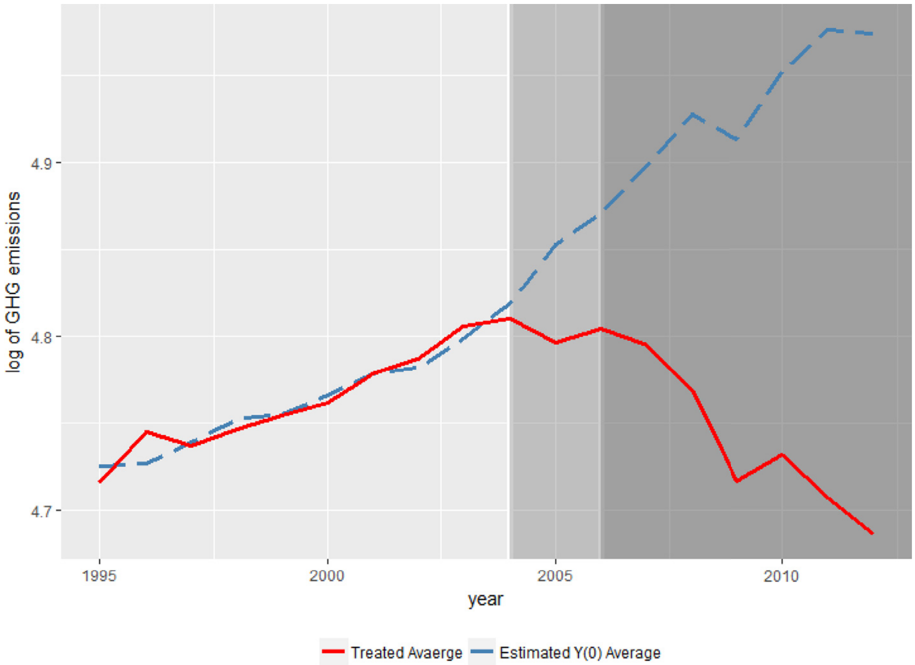

Fig. 7. The GHG emissions of the treated group and the counterfactual- FSU countries excluded from treated group. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies.

4.2.3. *Removing CDM host countries*

KP allowed for some flexibility mechanisms, such as CDM, where the Annex B countries were allowed to invest in emission reduction projects in the non annex B countries and then acquire more emission allowance. This of course fortifies the criticism of the differentiated targets between the industrialized (Annex B) countries and the developing countries. As well as poses as an empirical challenge as the control group may actually also be "treated" as a result of the protocol.[17] Conceptually the results should hold as the GSCM estimator formulates the counterfactual so that it is as similar as possible to the treated unit/group in the years before treatment. Given that the top host countries of CDM projects are China, India, Brazil, Mexico, Vietnam, Malaysia and Indonesia,[18] the probability that they constitute a large enough part of the counterfactual to drive the emissions significantly is very low given the differences between those countries and the treated group.[19] And while the driving concern for the CDM recipients is that they were also "treated" as donor (Annex B) countries invested in emission reductions there, the top recipients have also exhibited a rise in their GHG emissions as shown in Fig. 1 (China and east Asia and Pacific). In both cases their may be a bias in the results due to their presence in the control group, either: downward (due to emission reduction from CDM, in the sense that the gap between treated and counterfactual is underestimated) or upward (due to the rapid increase in the emissions of those countries leading to an overestimation of treatment effect).

Consequently, to check if they are actually biasing the emissions of the counterfactual, the top 5 CDM host countries that represent around 80% of the projects by 2012 are removed from the control group (China, India, Brazil, Mexico, Vietnam).

Results presented in the CDM column in Table 2 show an 8.5% reduction on GHG emissions that is highly significant (illustrated by Figs. 8 and 9). The results seem to be consistent (if somewhat higher) with the main results. This shows that the magnitude of the reduction of GHG emissions of the treated group in comparison with the counterfactual -shown in the main results I- is not driven by an upward bias in the counterfactual's emissions due to the presence of one of those countries in the control group. In fact, this result shows that their removal led to a higher treatment effect. This is in line with the first concern, that their presence in the control group had led to the underestimation of the treatment effect, though only slightly.

4.2.4. *Potential endogeneity: non-random treatment*

One of the main challenges that face empirical studies on international agreements is endogeneity arising from the voluntary selection into the agreement and the KP is no exception. The endogeneity in this case may arise from the fact that the countries

[17] In the sense that the countries with legal obligation would invest in emission reduction in the countries with no commitments affecting the emissions of those countries, which would not have occurred if they had continued BAU.

[18] Information on CDM projects from: http://cdm.unfccc.int/Statistics/Public/CDMinsights/index.html#reg.

[19] This does not mean that they would not be totally excluded from the counterfactual, however, their assigned weights are not expected to be high (enough to drive the results) given the difference between them and the treated group in the pre-treatment period.

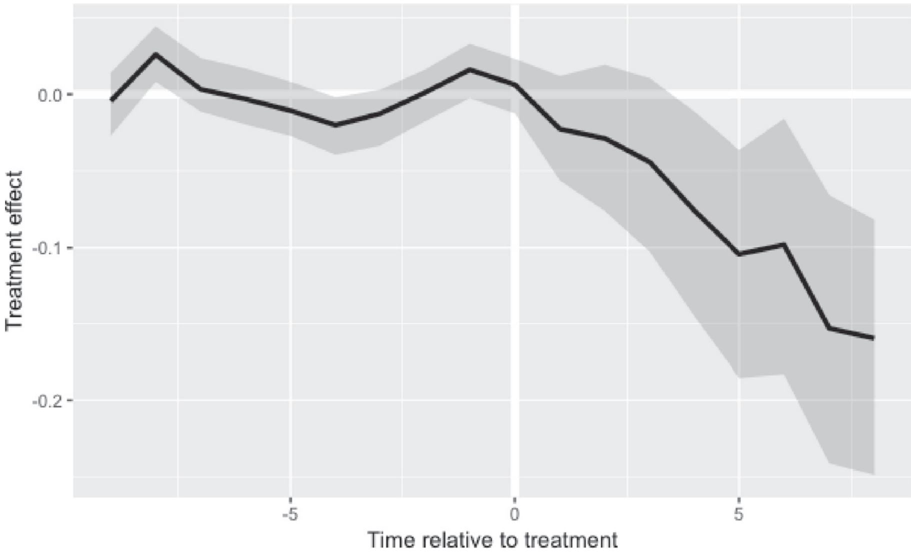

Fig. 8. The gap in the GHG emissions between the treated group and the counterfactual- CDM countries excluded from control group. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

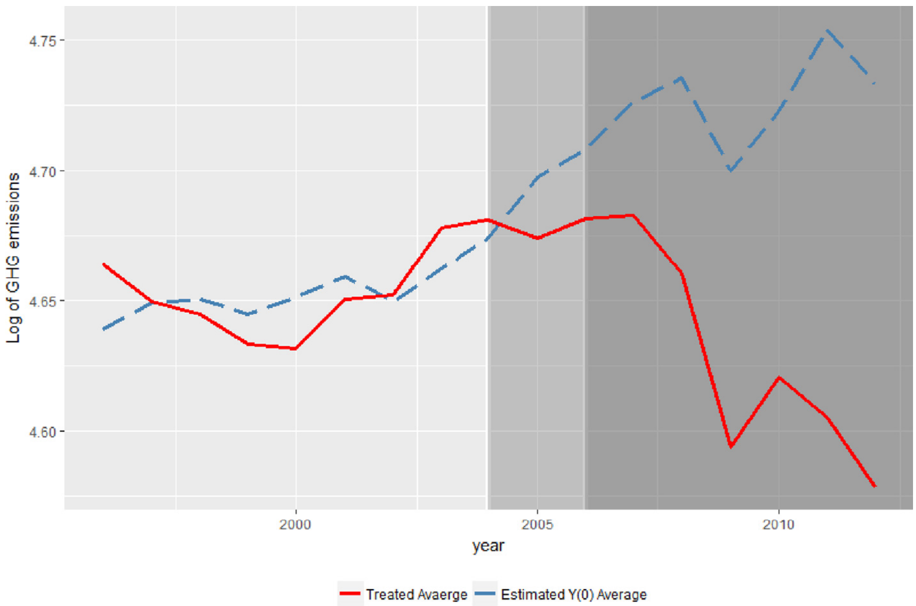

Fig. 9. The GHG emissions of the treated group and the counterfactual- CDM countries excluded from control group. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies.

that are legally bound by a target level were chosen by the UNFCCC based on their previous industrialization (pre-selection bias), so the treatment is not random. In addition to that not all those countries mentioned in Annex B ratified the protocol (self-selection), so selection is a probable issue. The use of the synthetic counterfactual reduces the bias as the unobservable confounders that may cause endogeneity should- at least theoretically- affect the counterfactual as well[20] (Billmeier and Nannicini, 2013), however, it does not eliminate the bias completely. The unobservable confounders found by the GSCM estimators should also alleviate the concern for endogeneity[21] as it should capture the inherent differences that could result

[20] Given that it formulates a synthetic country/group that is almost identical to the treated one. In this case, if the treated countries exhibit high emission levels previous to treatment then the synthetic counterfactual should also exhibit similar levels of emissions.

[21] Xu provides a detailed explanation for how the GSCM estimator should alleviate concerns for endogeneity in section 5 in his paper.

Table 3
The effect of Kyoto protocol on GHG emissions- using more control variables.

Variable	34 treated countries ATT/Beta coefficient
Ln GHG (excluding LUCF)	−0.072*** (0.034)
Ln Real GDP per capita	0.375*** (0.045)
Population Growth squared	−0.010* (0.0067)
GDP growth	−0.0004 (0.0005)
Human capital	0.328*** (0.126)
Life Expectancy	0.0128*** (0.0034)
Agriculture value added (% of GDP)	−0.0005 (0.0018)
Industry value added(% of GDP)	−0.0067* (0.0024)
Service value added(% of GDP)	3.13e-06 (0.00002)
Polity2	0.0014 (0.0016)
Political Rights	0.0033 (0.0050)
MSPE	0.00194
Unobserved factors	1
Treated units	33[a]
Control units	95

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *MSPE* is mean squared predictor error. Years of analysis 1995 till 2012. Treatment year is 2005, except Australia it is 2007. The treated countries are found in the 1st column in Table A.1 and the list of control countries is available in A.2.
[a] Iceland was dropped due to missing data.

from the treatment being non-random and since the GSCM estimator takes the unobservable confounders found into account when formulating the counterfactual, the concern for endogeneity should be eased (Xu, 2017).

Nonetheless, to ensure the robustness of the results, the potential endogeneity is addressed in several different ways. First, the industrialization factor should be taken into consideration as it is the base on which countries were chosen by the UNFCCC to be legally bound by the emission reduction targets. To account for the industrialization effect on the ratification of Annex B countries more covariates are used that would be reflective of the differences that may affect GHG emissions between the treated group and the control group. In their paper Almer and Winkler (2017) use a set of predictors to account for the underlying variation in GHG across countries such as economic power, industry structure and institutional quality. These predictors should particularly target the selection bias for treatment, given the inherent differences between the 2 groups of countries that might affect their GHG emissions. The second way to address it is by using the U.S. states as a control group as done in Almer and Winkler (2017) (in the following section 4.2.5), as they present a more comparable group to the Annex B countries.

Results show a 7.2% reduction in GHG emissions (presented in Table 3), which is consistent with the main results that showed an approximate of 7% reduction. The estimated counterfactual shown in Figs. 10 and 11 is also very similar to the ones estimated in the main results model and illustrated by Figs. 2 and 3. The similar and almost equivalent treatment effects in the two models is evidence of the robustness of the estimated counterfactual and provides evidence that the non-random selection of treatment did not prove to bias the estimated treatment effects (see Fig. 12–15).

4.2.5. Using the U.S. as control group

Another way to address the problem of the presence of potential unobserved differences -that arise due to the heterogeneous pool of controls- is to use the U.S. states as a control group as done in Almer and Winkler (2017). The advantage of using the U.S. as a control group lies in it being initially a part of the annex B countries and hence the selection criteria (of the UNFCCC) that the countries with commitments fulfilled, were also fulfilled by the U.S. and accordingly this reduces the pre-selection bias. Following Almer and Winkler (2017) the treated group consists of 15 countries (see Appendix A) as well as the same predictors[22]

[22] The same variables were chosen (growth rates and standardized dependent variable) as the comparing states are obviously smaller and comparing them to countries using absolute values would not be accurate.

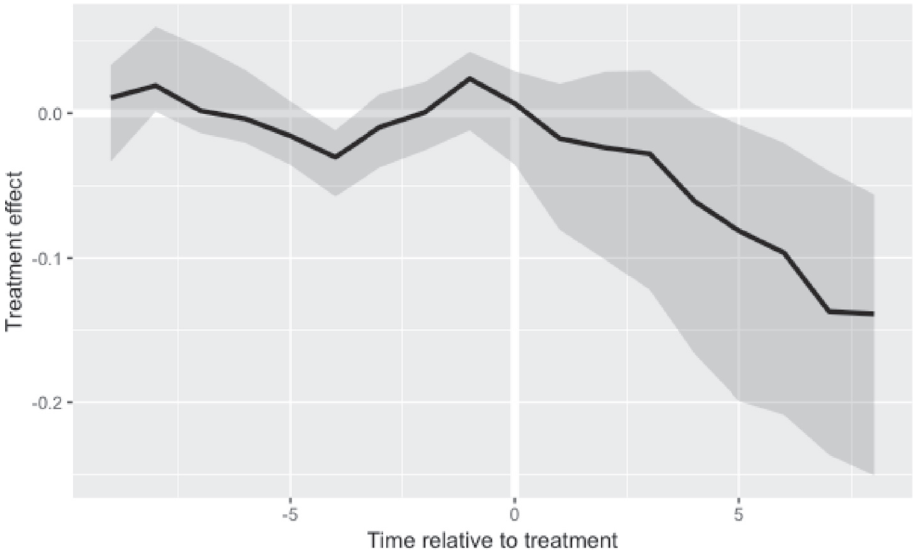

Fig. 10. The gap in the GHG emissions between the treated group and the counterfactual- using more control variables. The figure shows how the treated group's GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

Fig. 11. The GHG emissions of the treated group and the counterfactual- using more control variables. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies.

and the same dependent variable. The analysis period is from 1995 till 2011 instead 2012 as has been done throughout the paper due to the missing data on the dependent variable for the U.S..[23] Also, the treated group is limited to the same 15 countries that were analyzed by Almer and Winkler (2017) for two reasons. First, the 15 countries are mostly West and Northern European countries[24] which are more similar to the U.S. than some other countries included in Annex B (like FSU countries). Second, to make the comparison of results that is attributed to the use of the GSCM clearer.

[23] data on US states population and GDP are from Bureau of Economic Analysis (BEA) https://www.bea.gov/ and data on carbon dioxide is from CAIT Climate Data Explorer provided by World Resources Institute http://cait.wri.org.

[24] In addition to Australia, Canada and Japan.

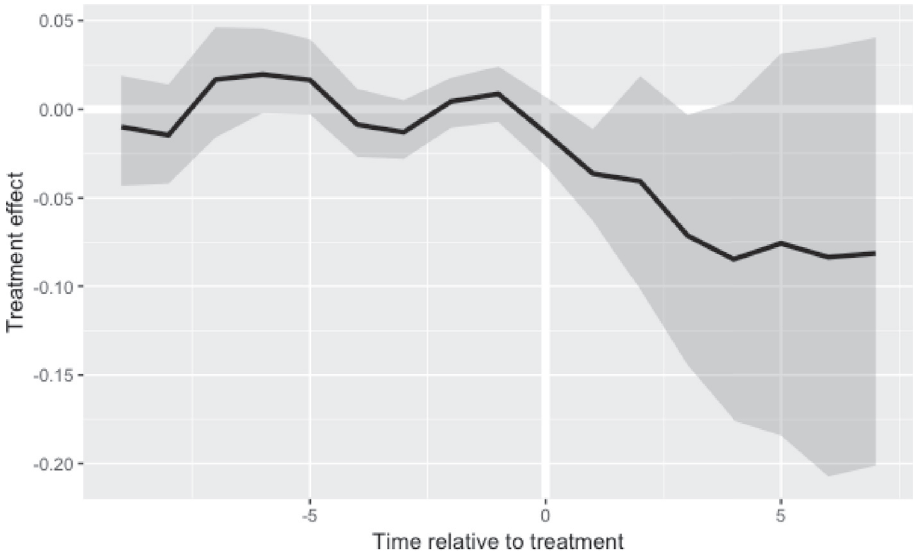

Fig. 12. The gap in the GHG emissions between the treated group and the counterfactual- U.S. states as control group. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval. The treated group here consists of 15 countries following Almer and Winkler [2017].

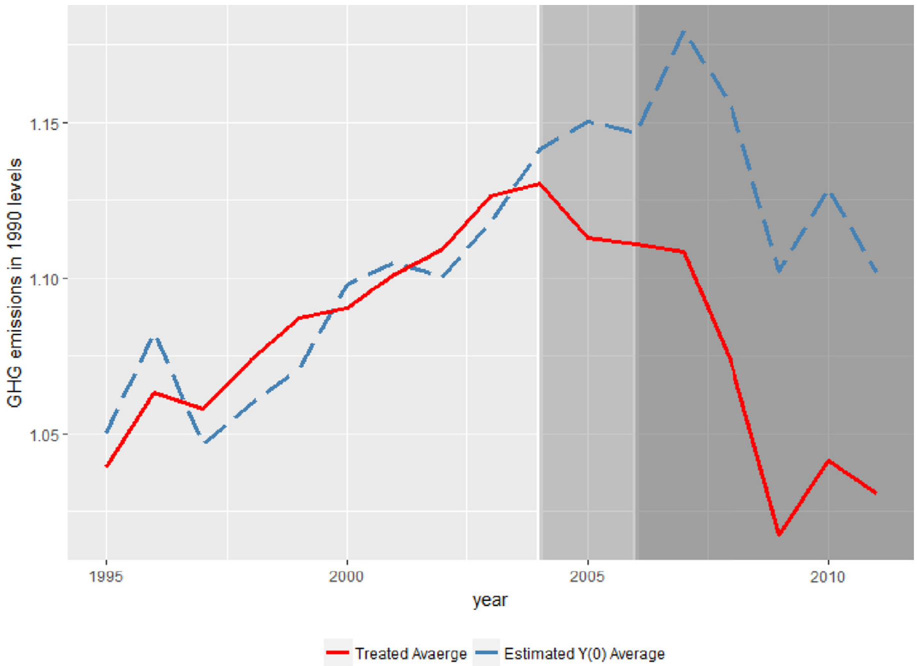

Fig. 13. The GHG emissions of the treated group and the counterfactual- U.S. states as control group. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies. The treated group here consists of 15 countries following Almer and Winkler [2017].

Both the GHG standardized by base year (1990) as well as CO_2 were used as dependent variable. The results presented in Table 4 show a 6.7% less emissions of GHG in 1990 levels of the treated group relative to the U.S. emissions.[25] As for the CO_2 emissions levels the reduction is around 7.4% significant on 5% level.[26] The second result con-

[25] Significant only on a 10% level.

[26] The slightly higher reduction in carbon dioxide emissions is explained by the fact that it constitutes the larger share of emitted GHG and hence when aiming to reduce GHG it would also constitute a larger share of the reduction.

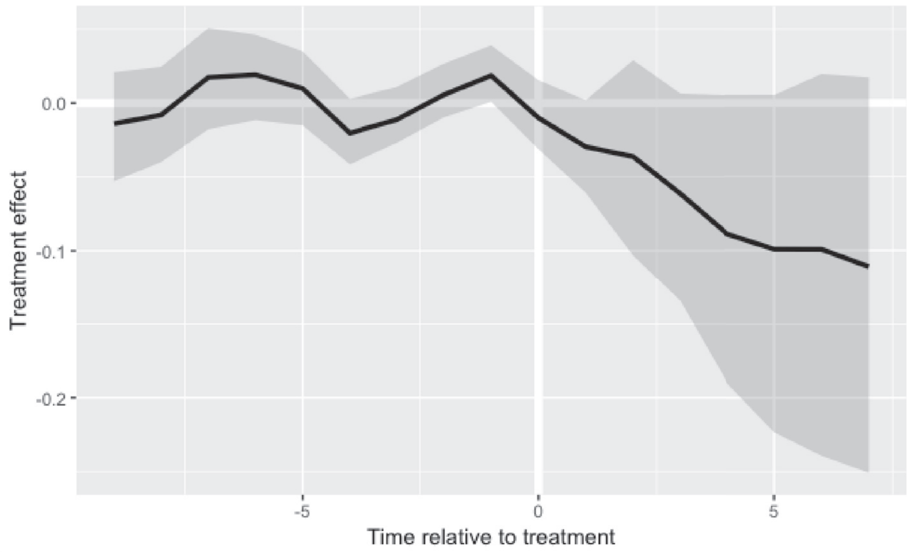

Fig. 14. The gap between CO_2 emissions in the treated group and the counterfactual- U.S. states as control group. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval. The treated group here consists of 15 countries following Almer and Winkler [2017].

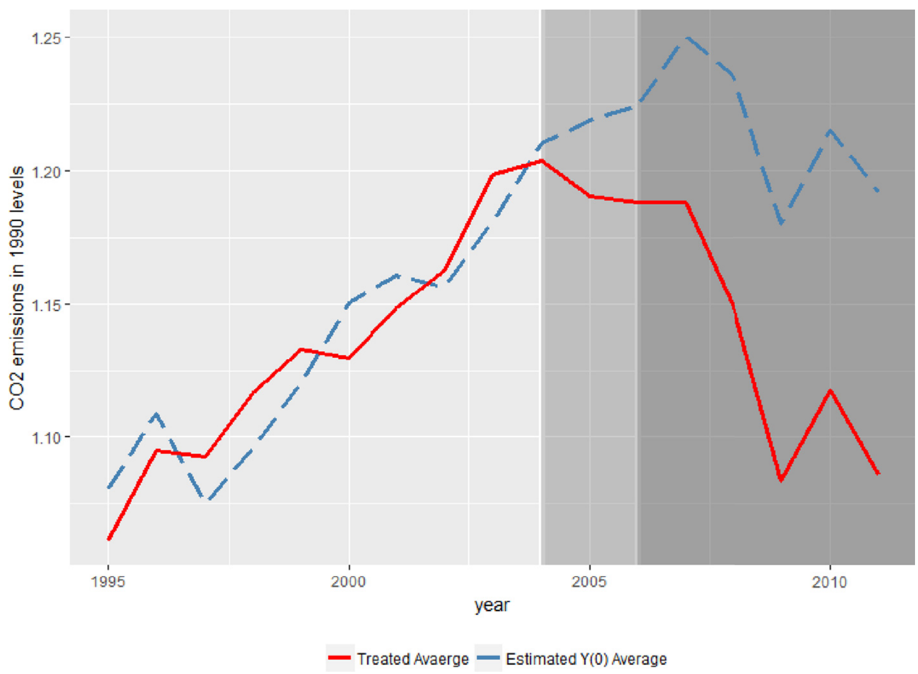

Fig. 15. The CO_2 emissions of the treated group and the counterfactual- U.S. states as control group. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies. The treated group here consists of 15 countries following Almer and Winkler [2017].

trasts with the results shown by Almer and Winkler (2017) of non-significant reduction in the CO_2 emissions of the treated group. The results seem to be consistent with the main result, which is evidence if the robustness of the GSCM estimator in formulation of the counterfactual and its emissions notwithstanding the differences in the control groups.

Table 4
The effect of KP on emission levels of treated countries- using U.S. states as control group.

Variable	U.S. as control group	
	GHG	CO_2
ATT	−0.067* (0.04)	−0.074** (0.04)
Ln Real GDP per capita	0.196* (0.09)	0.200* (0.10)
Population growth	0.014** (0.005)	0.014** (0.0006)
GDP growth	−0.0003 (0.0008)	−0.0005 (0.0008)
MSPE	0.00162	0.00227
Unobserved factors	1	1
Treated units	15	15
Control units	51	51

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *MSPE* is mean squared predictor error. Years of analysis 1995 till 2012. Treatment year is 2005, except Australia it is 2007. The treated countries are found in the 4th column in Table A.1 and the list of states is found in A.2.

5. Conclusion

International environmental agreements have been increasing in number over the past decades and so have the controversy surrounding their effectiveness. The Kyoto protocol, one of the most ambitious agreements tackling climate change problems at that time, was no exception to this debate. Its design has been criticized and its effectiveness has been questioned, and empirical evidence surrounding its effectiveness has provided controversial evidence.

One of the important characteristics of the Kyoto protocol is that it addressed the emission problem as a global problem that requires collective action and that was reflected in its design. The aim of this paper was to test effectiveness of the Kyoto protocol in achieving its main target of GHG emissions reduction using a newly introduced method (GSCM). By using the GSCM the emissions of the industrialized countries after ratification were compared with a synthetic counterfactual that represents a hypothetical scenario of the would-have-been emissions in the case of a "No-Kyoto" scenario. The GSCM facilitates a robust comparison between treated and control countries as done by Almer and Winkler (2017), while simultaneously accounting for the collective nature of the protocol by computing the effects for the multiple treated units as done byGrunewald and Martinez-Zarzoso (2016). The results show that the industrialized countries that ratified the protocol had at least an approximate of 6–7% reduction in their emissions as compared to their expected emissions if they had proceeded without the protocol's targets in mind. The results remained consistent and robust through different checks. Removing the EU from the group as well as controlling for the selection bias showed similar reduction effects. The main exception to this result was the removal of the FSU countries, which showed an increase in the magnitude of the reduction reaching 17%, this increase is mainly attributed to their assignment of hot air. The consistency in the estimated reduction shows the robustness of the GSCM estimator in formulating an accurate counterfactual even with the changes in the control as well as treated group.

Based on the results presented in the paper the Kyoto protocol represents a successful first step that at least prevented a worse emission level from taking place. Further steps are needed of course and this is realized by the global community as shown by the increasing number of agreements and treaties that are still taking place after the Kyoto protocol's commitment period ended. This also means that hope for the Paris agreement should not be lost after the de-ratification of the U.S.. As was shown in the case of the Kyoto protocol, the non-participation of some countries such as the U.S., did not render the treaty useless.

Conflict of interest statement

I certify that I have no affiliations with or involvement in any organization or entity with any financial interest (such as grants), or non-financial interest (such as personal affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Financial disclosure statement

I also certify that no party has a direct interest in the results of the research supporting this article or will confer a benefit on me or on any organization with which I am associated and, if applicable, I certify that all financial and material support for this research and work are clearly identified in the title page of the manuscript.

A. List of countries

A.1. Treated groups

Main	RC: EU	RC: FSU	RC: US states as control
Australia	Australia	Australia	Australia
Austria	–	Austria	Austria
Bulgaria	–	–	–
Belgium	–	Belgium	Belgium
–	–	–	Canada
Croatia	Croatia[a]	Croatia	–
Czech Republic	–	–	–
Denmark	–	Denmark	–
Estonia	–	–	–
Finland	–	Finland	Finland
France	–	France	France
Germany	–	Germany	Germany
Greece	–	Greece	–
Hungary	–	–	–
Iceland	Iceland	Iceland	–
Ireland	–	Ireland	–
Italy	–	Italy	Italy
Japan	Japan	Japan	Japan
Latvia	–	–	–
Lithuania	–	–	–
Luxembourg	–	Luxembourg	–
Netherlands	–	Netherlands	Netherlands
New Zealand	New Zealand	New Zealand	–
Norway	Norway	Norway	Norway
Poland	–	–	–
Portugal	–	Portugal	Portugal
Romania	–	–	–
Russia	Russia	–	–
Slovakia	–	–	–
Slovenia	–	–	–
Spain	–	Spain	Spain
Sweden	–	Sweden	Sweden
Switzerland	Switzerland	Switzerland	–
United Kingdom	–	United Kingdom	United Kingdom
Ukraine	Ukraine	–	–
Total: 34	9	22	15

Main refers to the list of treated countries used in main results as well as the *CDM* in 4.2.3. *RC* is robustness check. *RC:EU* is the list of treated countries in the robustness check 4.2.1, excludes EU countries. *RC: FSU* is the list of treated countries in the robustness check 4.2.2, excludes former soviet union (FSU) countries. *RC: US as control* is the list of treated countries in 4.2.5, following Almer and Winkler (2017).

[a] Joined EU in 2013 after the end of the first commitment period.

A.2. Control groups

119 countries					
Albania	Algeria	Angola	Argentina	Armenia	Azerbaijan
Bangladesh	Belarus	Benin	Bhutan	Bolivia	Bosnia & Herzegovina
Botswana	Brazil	Burkina Faso	Burundi		
Cambodia	Cameroon	Canada	Central African Republic	Chad	Chile
China	Colombia	Comoros	Congo, Dem. Rep.	Congo, Rep.	Costa Rica
Cote d'Ivoire	Cuba	Cyprus	Djibouti	Dominica	Dominican Rep.
Ecuador	Egypt	El Salvador	Equatorial Guinea	Ethiopia	Fiji
Gabon	Gambia	Georgia	Ghana	Guatemala	Guinea
Guinea-Bissau	Guyana	Honduras	India	Indonesia	Iran, Islamic Rep.
Iraq	Israel	Jamaica	Jordan	Kazakhstan	Kenya
Korea, Rep.	Kuwait	Kyrgyzstan	Lao PDR	Lebanon	
Lesotho	Liberia	Macedonia	Madagascar	Malawi	Malaysia
Mali	Mauritania	Mauritius	Mexico	Moldova	Mongolia
Morocco	Mozambique	Myanmar	Namibia	Nepal	Nicaragua
Niger	Nigeria	Oman	Pakistan	Panama	Papua New Guinea
Paraguay	Peru	Philippines	Rwanda	Saudi Arabia	Senegal
Serbia	Sierra Leone	Singapore	South Africa	Sri Lanka	Sudan
Suriname	Swaziland	Tajikistan	Tanzania	Thailand	Togo
Trinidad & Tobago	Tunisia	Turkey	Turkmenistan	Uganda	United Arab Emirates
United States	Uruguay	Uzbekistan	Venezuela	Vietnam	Yemen
Zambia	Zimbabwe				

119 control countries used in the main results as well as the robustness checks for the EU 4.2.1, Hot air 4.2.2 and in appendix B.2. They were also used for the *CDM* robustness check 4.2.3 where top CDM host countries were removed (China, India, Brazil, Mexico and Vietnam), ending up with 114 countries.

51 states					
Alabama	Alaska	Arizona	Arkansas	California	Colorado
Connecticut	Delaware	District of Columbia	Florida	Georgia	Hawaii
Idaho	Illinois	Indiana	Iowa	Kansas	Kentucky
Louisiana	Maine	Maryland	Massachusetts	Michigan	Minnesota
Mississippi	Missouri	Montana	Nebraska	Nevada	New Hampshire
New Jersey	New Mexico	New York	North Carolina	North Dakota	Ohio
Oklahoma	Oregon	Pennsylvania	Rhode Island	South Carolina	South Dakota
Tennessee	Texas	Utah	Vermont	Virginia	Washington
West Virginia	Wisconsin	Wyoming			

51 U.S. states (50 states and District of Colombia) used as control group in the robustness check 4.2.5.

95 countries					
Albania	Algeria	Angola	Argentina	Armenia	Bangladesh
Benin	Bolivia	Botswana	Brazil	Burkina Faso	Burundi
Cambodia	Cameroon	Canada	Central African Republic	Chile	China
Colombia	Congo	Costa Rica	Côte d'Ivoire	Cyprus	Congo, Dem. Rep.
Dominican Rep.	Ecuador	Egypt	El Salvador	Ethiopia	Fiji
Gabon	Gambia	Ghana	Guatemala	Honduras	India
Indonesia	Iran	Israel	Jamaica	Jordan	Kazakhstan
Kenya	Kuwait	Kyrgyzstan	Lao PDR	Lesotho	Liberia
Madagascar	Malawi	Malaysia	Mali	Mauritania	Mauritius
Mexico	Mongolia	Morocco	Mozambique	Myanmar	Namibia
Nepal	Nicaragua	Niger	Nigeria	Pakistan	Panama
Paraguay	Peru	Philippines	Qatar	Korea, Rep.	Moldova
Rwanda	Saudi Arabia	Senegal	Sierra Leone	Singapore	South Africa
Sri Lanka	Swaziland	Syria	Tajikistan	Thailand	Togo
Trinidad & Tobago	Tunisia	Turkey	Uganda	United Arab Emirates	United States
Uruguay	Venezuela	Vietnam	Yemen	Zambia	

95 countries used as control group in the robustness check using more control variables 4.2.4 and in the Appendix B.1.

B. More robustness checks

B.1. Selection bias: More predictors

To make the results more comparable to the paper by Almer and Winkler (2017) that shows that the KP had an insignificant effect on the committed countries and to show also that the results shown in section 4.2.4 are not biased due to the set of countries used, the 15 treated countries used by Almer and Winkler (2017) were used. Due to missing data for Canada in one of the predictors it was removed.

Results show a 10.4% reduction highly significant (see Table 5). The results seem to be consistent with previous results in that the removal of the other 19 countries seem to lead to a bigger reduction and it could be driven by the FSU as explained in section 4.2.2.

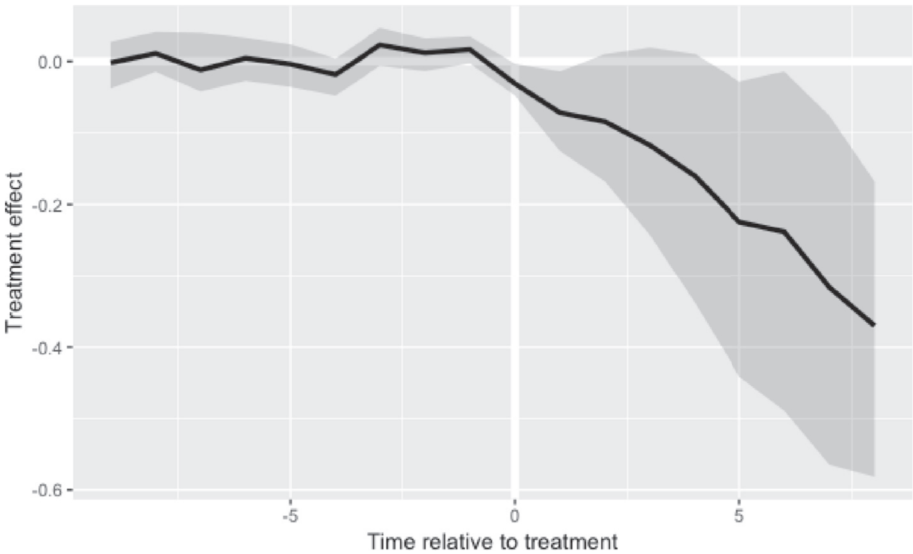

Fig. 16. The gap in the GHG emissions between the treated group and the counterfactual- using more control variables. The figure shows how the treated group's GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval. The treated group here consists of 15 countries following Almer and Winkler [2017].

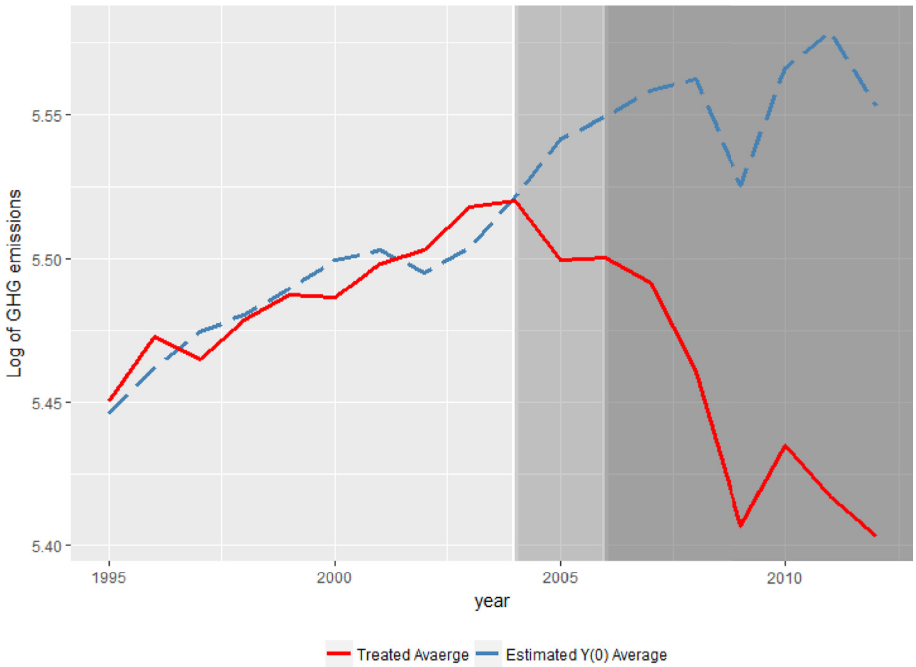

Fig. 17. The GHG emissions of the treated group and the counterfactual- using more control variables. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies. The treated group here consists of 15 countries following Almer and Winkler [2017].

Table 5
The effect of Kyoto protocol on GHG emissions-using more control variables.

Variable	ATT/Beta coefficient
Ln GHG (excluding LUCF)	−0.104***
	(0.033)
Ln Real GDP per capita	0.384***
	(0.041)
Population Growth squared	−0.010*
	(0.0054)
GDP growth	−0.0004
	(0.0005)
Human capital	0.238***
	(0.127)
Life Expectancy	0.0120***
	(0.0032)
Agriculture value added (% of GDP)	−0.0001
	(0.0016)
Industry value added(% of GDP)	−0.0062**
	(0.0023)
Service value added(% of GDP)	5.71e-06
	(0.00002)
Polity2	0.0010
	(0.0016)
Political Rights	0.0029
	(0.0047)
MSPE	0.00116
Unobserved factors	1
Treated units	14[a]
Control units	94

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *MSPE* is mean squared predictor error. Years of analysis 1995 till 2012. Treatment year is 2005, except Australia it is 2007. The treated countries are found in the 4th column in Table A.1 and the control countries in A.2.

[a] Canada was dropped due to missing data.

B.2. Removing the squared GDP per capita

The debate as to whether there exists an Environmental Kuznets Curve for GHG and in specifically CO_2 has provided controversial evidence and hence the existence of a non-linear relationship between income and emissions might be questionable (Ansuategi and Escapa, 2002; Cho et al., 2014). To make the results more comprehensive, the squared GDP per capita term was removed, to check whether its removal will lead to a change of results. Removing the squared term of the real GDP per capita, shows the result to be exactly the same as the those provided in the main results as well as the relevant robustness checks. A 6.8% reduction in emissions is shown in Table 6.

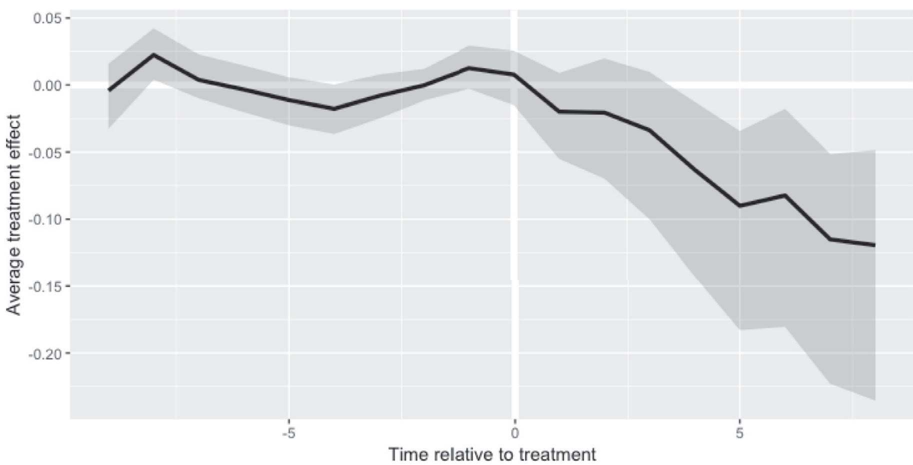

Fig. 18. The gap in the GHG emissions between the treated group and the counterfactual- removing the squared GDP per capita from the control variables. The figure shows how the treated group's GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

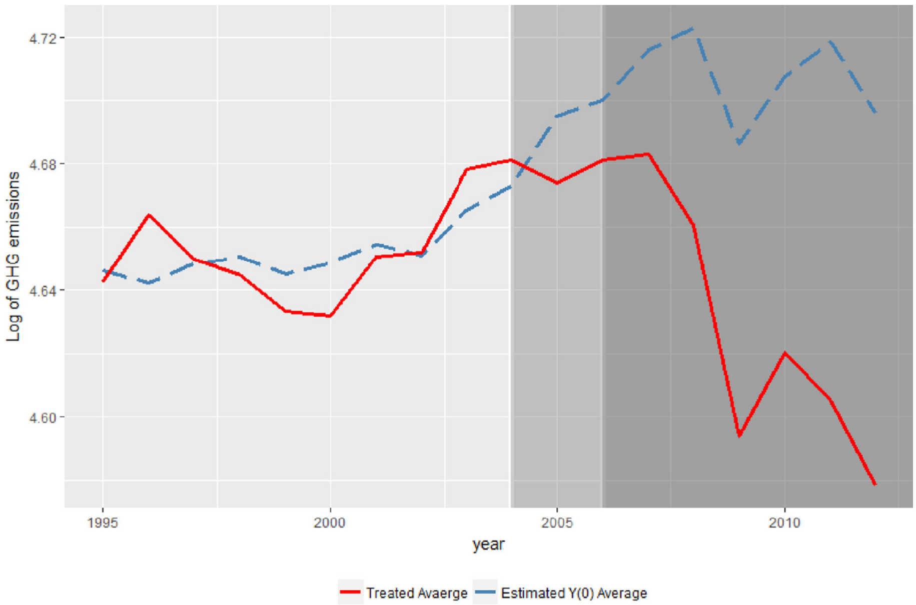

Fig. 19. The GHG emissions of the treated group and counterfactual- removing the squared GDP per capita from the control variables. The red solid line represents the GHG emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2005) is shaded in a darker grey, where it becomes even darker post 2007 when Australia ratifies.

Table 6
The effect of Kyoto protocol on GHG emissions on industrialized countries - removing the squared GDP per capita from the control variables.

Variable	ATT/Beta coefficient
Ln GHG (excluding LUCF)	−0.067*** (0.030)
Ln Real GDP per capita	0.38*** (0.288)
Ln population	0.736*** (0.160)
MSPE	0.00159
Unobserved factors	1
Treated units	34
Control units	119

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *LUCF* is land-use change and forestry. *MSPE* is mean squared predictor error. Years of analysis 1995 till 2012. Treatment year is 2005, except Australia it is 2007. The treated countries are found in the 1st column in Table A.1 and the control countries in A.2.

C. Replications

C.1. A similar analysis to Almer and Winkler (2017)

In this specification, the treated group consists of 15 countries as identified by Almer and Winkler (2017), the control group consists of high and upper middle income countries as characterized by the world bank in 2002. The years of analysis cover from 1990 till 2012 with 2002 as the treatment year and the dependent variables are GHG as 1990 levels and CO_2 as 1990 levels (in natural logs). This is to increase comparability to the specification used in Almer and Winkler (2017).

Results show an approximate 6% and 8% reduction in GHG and CO_2, respectively. The estimated reduction is similar although slightly lower in magnitude to the emission reduction in the main specification, however it is not statistically significant on a 5%

level. Also the control group is only 34 countries (less than the recommended 40 units). Accordingly, results should be regarded carefully and conclusions should be made with caution.

One interesting observation with regards to the results is that the divergence between the treated and counterfactual in both cases (GHG and CO_2) takes place a few years (2–3 years) after the assigned treatment year (2002) as shown in Figs. 20–23. This shows that even if the method formulates the counterfactual emissions to be similar in the years up to 2002, the actual divergence in the emissions takes place after the year 2005. This result may point to several interesting conclusions. On the one hand, it could explain why there is still a difference between the results presented in this paper and the results presented in Almer and Winkler (2017); in addition to accounting for the collective nature of the protocol, the choice of treatment year could be one of the reasons why in this paper, when the enforcement year (2005) was chosen for treatment, the post KP emissions of the treated group was shown to be significantly different from the counterfactual, as compared to the when the ratification year (2002) was set as treatment year. A similar observation in the time of divergence is shown in the next section C.2. There are several reasons why in the case of KP, the year of enforcement is a more reasonable choice for treatment.

- Australia did not ratify until 2007.
- While the hype (politically) usually takes place around the ratification period domestically as pointed out by several studies in the literature (Aichele and Felbermayr, 2012; Grunewald and Martinez-Zarzoso, 2016), however, in the case of KP the target reductions were not annual rather per commitment period (2008–2012), which gave countries some time to act. Also as mentioned in the protocol in article 7, the protocol requires countries to communicate their emissions only after it enters into force (United Nations, 1998).
- And even if countries do set new policies domestically at the exact time of ratification, it is reasonable to think that it may take time between putting these policies in practice and seeing actual reduction in GHG. So a noticeable effect on emissions may not be seen directly after the ratification period starts.
- KP's commitment period does not begin till 2008, so 2005 made sense as a middle ground (between ratification and commitment).

One main limitation of this result is that the control group is smaller than 40 units, which is the minimum recommended by Xu (2017), as a small control group may cause a bias in the result due to excessive extrapolation.

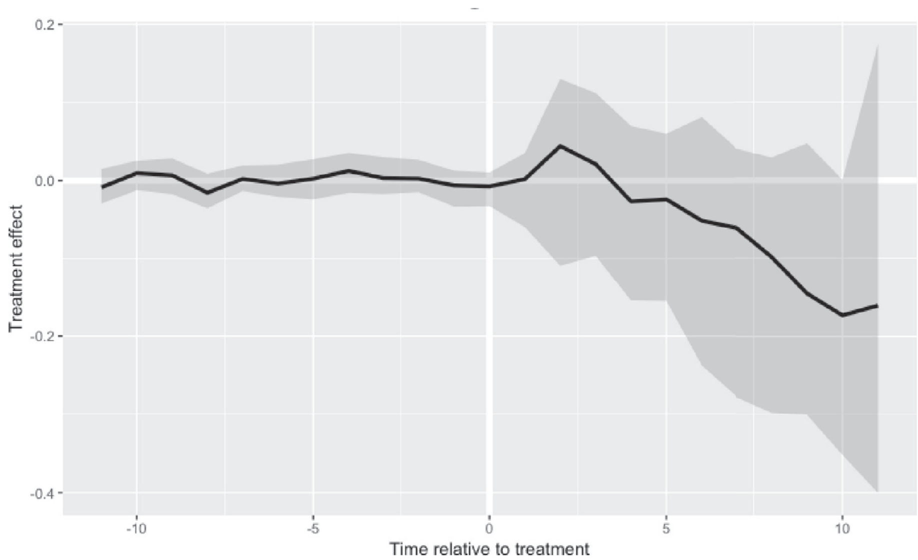

Fig. 20. Gap between the GHG emissions of the treated group and the counterfactual. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

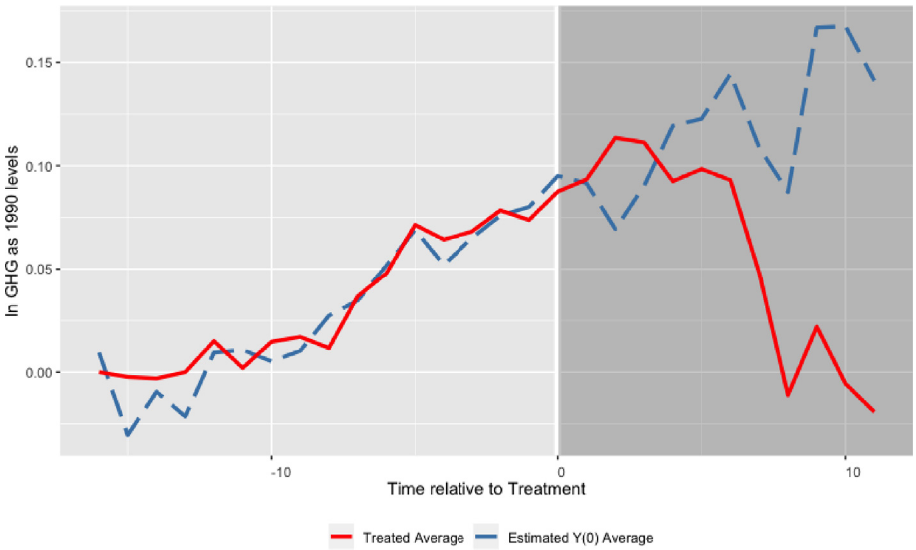

Fig. 21. GHG emissions of the treated group and the counterfactual. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2002) is shaded in a darker grey.

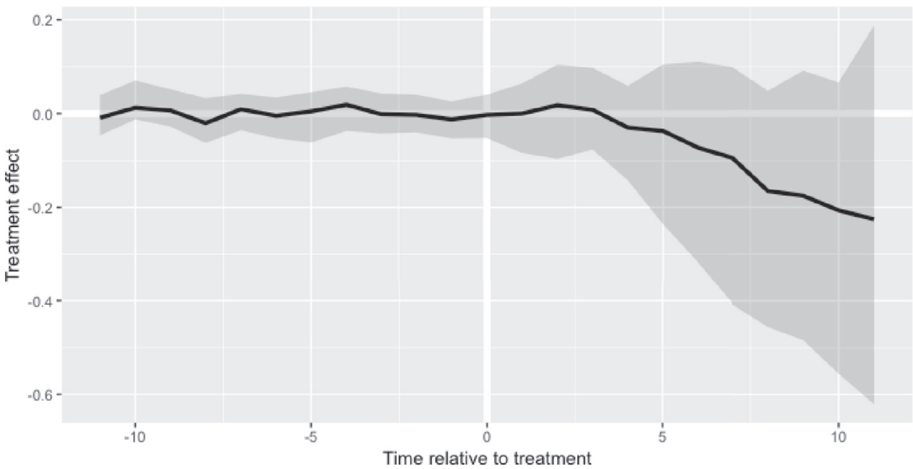

Fig. 22. Gap between CO_2 emissions of the treated group and the counterfactual. The figure shows how the treated group's average CO_2 emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the CO_2 emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

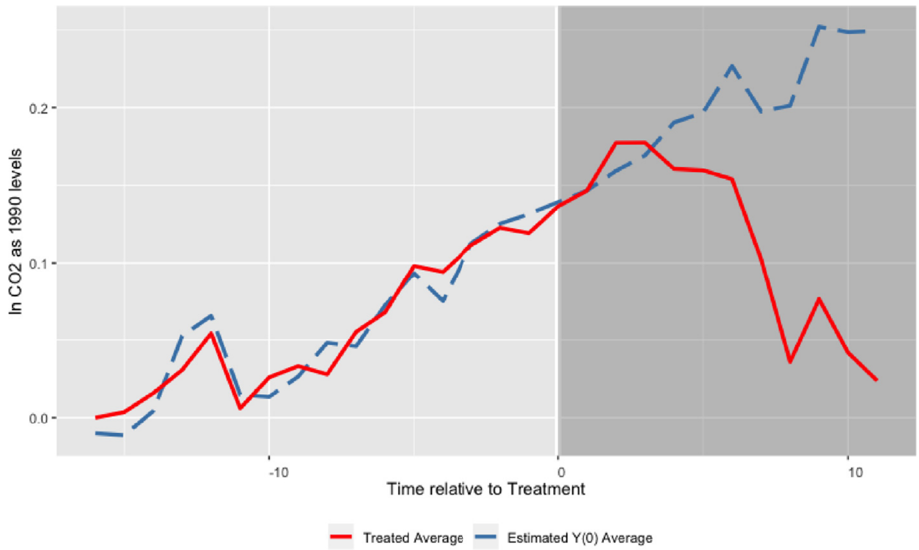

Fig. 23. CO_2 emissions of the treated group and the counterfactual. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2002) is shaded in a darker grey.

Table 7
The effect of KP on emission levels of treated countries.

Variable	*GHG*	CO_2
ATT	−0.059	−0.086
	(0.05)	(0.09)
Ln Real GDP per capita	−1.47e+04	−11861.7
	(1.67e+04)	(2.53e+04)
Ln Population	0.93**	0.99**
	(0.28)	(0.40)
ln GDP per capita (squared)	7.36e+03	5931.17
	(8.39e+03)	(1.26e+04)
MSPE	0.00109	0.00145
Unobserved factors	3	1
Treated units	15	15
Control units	34	34

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *MSPE* is mean squared predictor error. Years of analysis 1990 till 2012. Treatment year is 2002, except Australia it is 2007. The list of treated countries is available in the 4th columns in appendix in the main manuscript. Control group consists of Antigua and Barbuda, Argentina, The Bahamas, Bahrain, Barbados, Botswana, Brazil, Brunei Darussalam, Chile, Costa Rica, Cyprus, Dominica, Gabon, Grenada, South Korea, Lebanon, Malaysia, Malta, Mauritius, Mexico, Oman, Panama, Saudi Arabia, Seychelles, Singapore, South Africa, St. Kitts and Nevis, St. Lucia, Trinidad and Tobago, Turkey, United Arab Emirates, United States, Uruguay and Venezuela.

C.2. A similar analysis to Grunewald and Martinez-Zarzoso (2016)

This specification aims at making the comparison between this paper and the paper by Grunewald and Martinez-Zarzoso (2016) more feasible. Although the difference between the specifications in both papers was not big, there were minor differences and this specification addresses them. The main differences were: the time period under analysis, the treatment year, the use of GHG instead of CO_2 and the number of countries. The time period analyzed by Grunewald and Martinez-Zarzoso (2016) was 1992–2009 and they used the ratification year for each country as the start of treatment year. With regards to the countries used, they identify 32 countries from Annex B and 138 control countries, yet the exact list is not provided in their paper. In this specification the period analyzed is similar to that of Grunewald and Martinez-Zarzoso (2016) and the ratification year for each country is used as a the start of treatment. Both GHG and CO_2 were used as dependent variable. The treated countries are 28 countries (listed in the Table 9) and the control countries are 110 (found in the footnotes of Table 8). The number of countries analyzed were less than that analyzed by Grunewald and Martinez-Zarzoso (2016), however, this could not be helped as the lack of data from 1992 till 1995 for some countries necessitated that they get dropped (a balanced dataset is necessary).

Results presented in table the 2nd column in Table 8 and illustrated in Figs. 24 and 25, show a similar emission reduction to that presented in the main results, a 6.5% reduction in the GHG emissions of the treated countries relative to the control group (significant at 5% level). In the 3rd column in Table 8, as well as Figs. 26 and 27, the ATT using CO_2 emissions is shown. The emission reduction is slightly higher ∼8%, however not statistically significant on a 5% level. The results are more or less consistent with those presented in Grunewald and Martinez-Zarzoso (2016), where they show at least a 7% reduction in emissions. They are also consistent with results in section 4.2.5 in the main manuscript and section C.1, where the reduction in CO_2 emissions was shown to be consistently slightly higher than that of GHG. The results are also consistent with the results presented in the previous section C.1, in that the divergence between the emissions of the treated group and the counterfactual takes place 2–3 years post the set treatment (in this case ratification years, which for most countries are around 2002). This is an interesting result in itself, that the countries' emissions start to decline around 2005 (the year of enforcement) rather than the year of ratification.

This estimated reduction is also in line with the results presented in Grunewald and Martinez-Zarzoso (2016), where they estimated a ∼ 7% reduction in CO_2 emissions. The similarity between the results presented in Grunewald and Martinez-Zarzoso (2016) and those presented in this section as well as in the main specification and the difference between them and the results presented in Almer and Winkler (2017) shows that accounting for the collective nature of the group may be the driving reason for the conflict in results.

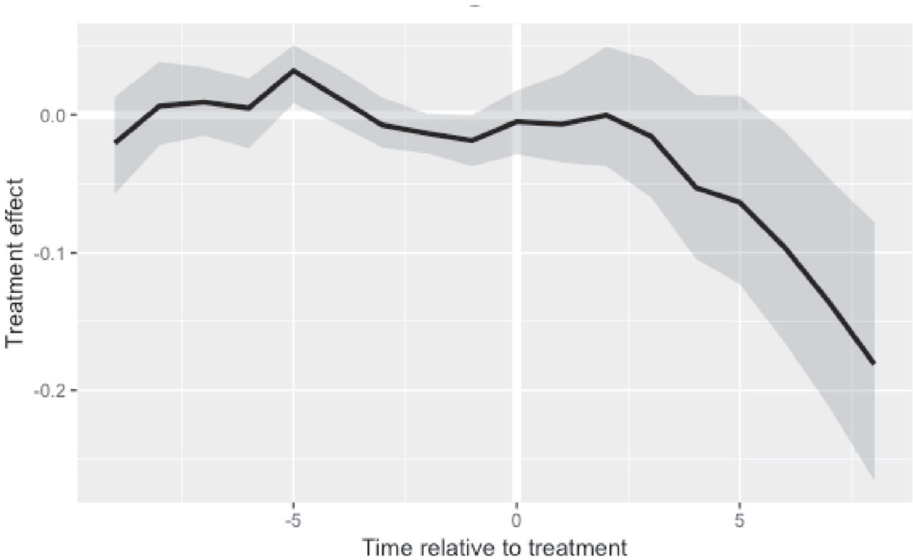

Fig. 24. Gap between the GHG emissions the treated group and the counterfactual. The figure shows how the treated group's average GHG emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the GHG emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

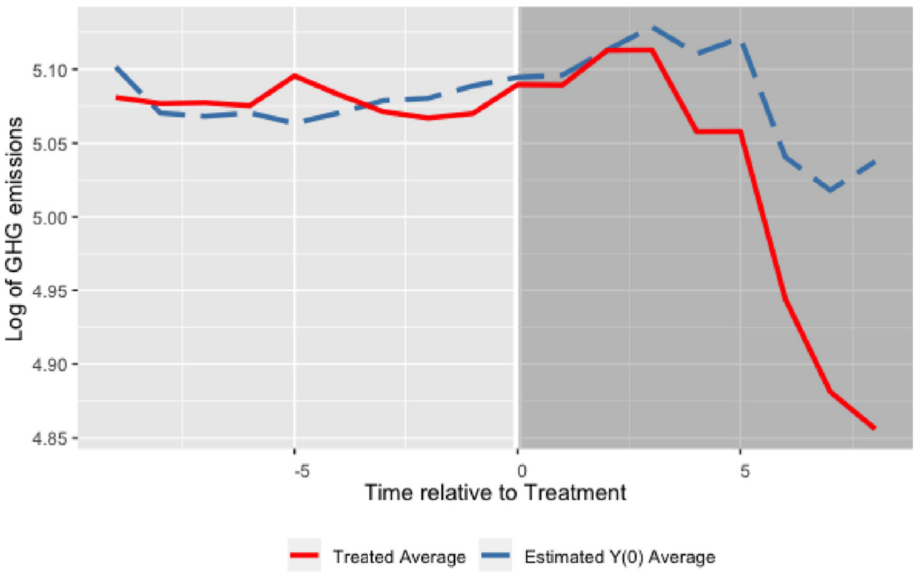

Fig. 25. GHG Emissions of the treated group and the counterfactual. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period is shaded in a darker grey.

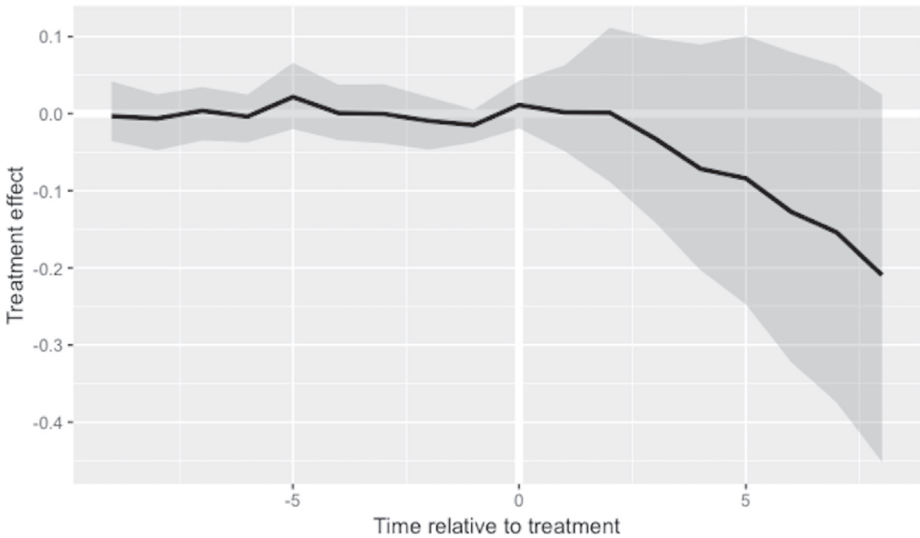

Fig. 26. Gap between CO_2 emissions of the treated group and the counterfactual. The figure shows how the treated group's average CO_2 emissions diverge from the counterfactual over time. The horizontal line (= zero) represents the counterfactual and the black line (downward sloping) represents the CO_2 emissions of the treated group. The vertical line (=0) is the treatment year. The grey shaded area is the 95% confidence interval.

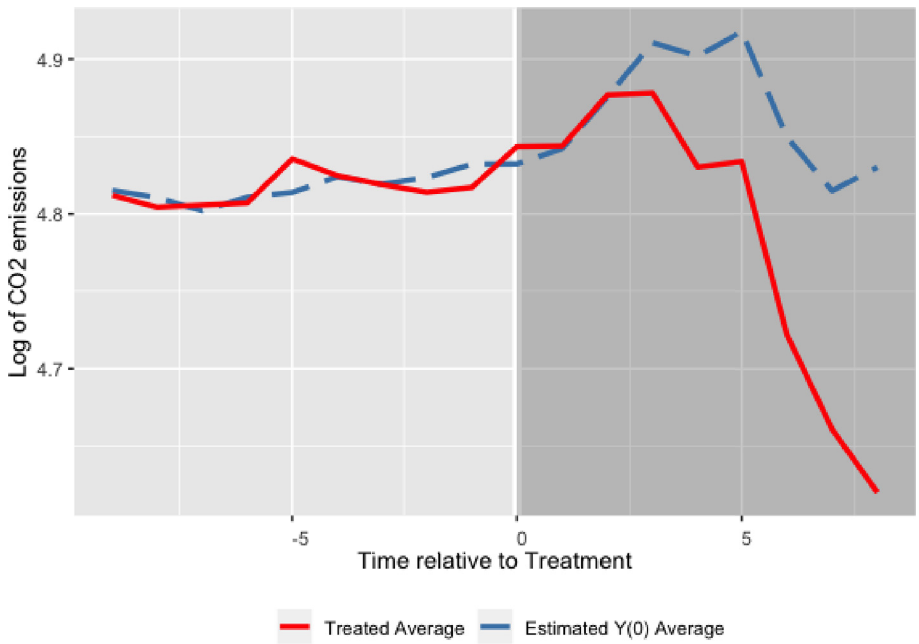

Fig. 27. CO_2 emissions of the treated group and the counterfactual. The red solid line represents the emissions of the treated group and the blue dashed line represents the emissions of the synthetic counterfactual. Treatment period (starting 2002) is shaded in a darker grey.

Table 8
The effect of KP on emission levels of treated countries (following a specification similar to Grunewald and Martinez-Zarzoso (2016)).

Variable	*GHG*	CO_2
ATT	−0.065**	−0.08
	(0.029)	(0.073)
Ln Real GDP per capita	−4663.5	−11578.9
	(2.07e+04)	(2.99e+04)
Ln Population	0.41**	1.0001**
	(0.15)	(0.39)
ln GDP per capita (squared)	2331.9	5789.5
	(1.03e+04)	(1.49)
MSPE	0.00277	0.00185
Unobserved factors	1	2
Treated units	28	28
Control units	110	110

*** significant 1% level, ** significant on 5% level, * significant on 10% level. Standard errors shown in the parentheses below each value. *MSPE* is mean squared predictor error. Years of analysis 1992 till 2009. Treatment year is the ratification year per country. Treated units and their corresponding ratification year available in the Table 9. Control units available in the 1st list in A.2 with the exception of Bosnia and Herzegovina, Cambodia, Djibouti, Dominica, Israel, Kuwait, Moldova, Serbia, Suriname, due to lack of data availability prior to 1995.

Table 9
Annex B and their corresponding ratification year.

Australia - 2007	Austria - 2002	Belgium-2002	Bulgaria –2002
Czech Republic - 2002	Denmark- 2002	Finland - 2002	France- 2002
Germany - 2002	Greece - 2002	Hungary- 2002	Ireland - 2002
Italy - 2002	Japan - 2002	Luxembourg - 2002	Netherlands - 2002
New Zealand - 2002	Norway - 2002	Poland - 2002	Portugal- 2002
Romania- 2002	Russian Federation –2005	Slovak Republic - 2002	Spain –2002
Sweden-2002	Switzerland –2003	United Kingdom - 2002	Ukraine-2004

Appendix D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jeem.2019.04.001.

References

Aakvik, A., Tjøtta, S., 2011. Do collective actions clear common air? The effect of international environmental protocols on sulphur emissions. Eur. J. Political Econ. 27 (2), 343–351.
Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J. Am. Stat. Assoc. 105 (490), 493–505.
Aichele, R., Felbermayr, G., 2012. Kyoto and the carbon footprint of nations. J. Environ. Econ. Manag. 63 (3), 336–354.
Aichele, R., Felbermayr, G., 2013. The effect of the Kyoto protocol on carbon emissions. J. Policy Anal. Manag. 32 (4), 731–757.
Aldy, J.E., Stavins, R.N., 2008. Climate policy architectures for the post Kyoto world. Environment 50 (3), 6–17.
Aldy, J.E., Stavins, R.N., 2013. Designing the PostKyoto climate regime. In: A New Global Covenant: Protection without Protectionism. Columbia University Press, New York, pp. 205–230.
Almer, C., Winkler, R., 2017. Analyzing the effectiveness of international environmental policies: the case of the Kyoto Protocol. J. Environ. Econ. Manag. 82, 125–151.
Ansuategi, A., Escapa, M., 2002. Economic Growth and Greenhouse Gas Emissions. Technical report.
Arce M, D.G., Sandler, T., 2001. Transnational public goods: strategies and institutions. Eur. J. Political Econ. 17, 493–516.
Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77 (4), 1229–1279.
Barrett, J., Andres, R., Bergesen, J., Gerlagh, R., Suh, S., Barrett, J., de Coninck, H.C., Diaz Morejon, C.F., Mathur, R., Nakicenovic, N., Ofosu Ahenkora, A., Pan, J., Pathak, H., Rice, J., Richels, R., Smith, S.J., Stern, D.I., Toth, F.L., Zhou, P., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., Minx, J., 2014. Drivers, trends and mitigation. In: Climate Change 2014: Mitigation of Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
Barrett, S., 1994. Self-enforcing international environmental agreements. Oxf. Econ. Pap. 46 (46), 878–894.
Barrett, S., 1998. Political economy of the Kyoto protocol. Oxf. Rev. Econ. Pol. 14 (4), 20–39.
Bennear, L.S., Coglianese, C., 2005. Measuring progress: program evaluation of environmental polices. Environment 47 (2), 22–39.
Billmeier, A., Nannicini, T., 2013. Assessing economic liberalization episodes: a synthetic control approach. Rev. Econ. Stat. 95 (3), 983–1001.
Böhringer, C., 2003. The Kyoto protocol: a review and perspectives. Oxf. Rev. Econ. Pol. 19 (3), 451–466.

Böhringer, C., Vogt, C., 2004. The dismantling of a breakthrough: the Kyoto protocol as symbolic policy. Eur. J. Political Econ. 20 (3), 597–617.
Breidenich, C., Magraw, D., Rowley, A., Rubin, J.W., 1997. The Kyoto protocol to the united nations framework convention on climate change. Am. J. Int. Law 92 (2), 315–331.
Cho, C., Chu, Y.P., Yang, H.Y., 2014. An environment Kuznets curve for GHG emissions: a panel cointegration analysis. Energy Sources Part B Energy Econ. Plan. Policy 9 (2), 120–129.
Faure, M., Lefevere, J., 2012. Compliance with Global Environmental Policy European Commission. The Global Environment Institutions, Law and Policy, pp. 163–180.
Feeny, D., Berkes, F., Mccay, B.J., Acheson, J.M., 1990. The tragedy of the commons: twenty-two years later. Hum. Ecol. 18 (1), 1–19.
Grossman, G.M., Krueger, A.B., 1995. Economic growth and the environment. Q. J. Econ. 110 (2), 353–377.
Grunewald, N., Martinez-Zarzoso, I., 2016. Did the Kyoto protocol fail? An evaluation of the effect of the Kyoto protocol on CO2 emissions. Environ. Dev. Econ. 21 (01), 1–22.
Helm, C., Sprinz, D.F., 2000. Measuring the effectiveness of international environmental regimes. J. Confl. Resolut. 44 (5), 630–652.
Kreif, N., Grieve, R., Hangartner, D., Turner, A.J., Nikolova, S., Sutton, M., 2016. Examination of the synthetic control method for evaluating health policies with multiple treated units. Health Econ. 25 (12), 1514–1528.
Libecap, G.D., 2014. Addressing global environmental externalities: transaction costs considerations. J. Econ. Lit. 52 (3), 424–479.
Livermore, M.A., Revesz, R.L., 2017. Environmental Law and economics. In: Oxford Handbook of Law and Economics, pp. 1–23.
Mickwitz, P., 2003. A framework for evaluating environmental policy instruments. Evaluation 9 (4), 415–436.
Mitchell, R.B., 2003. International environmental agreements: a survey of their features, formation, and effects. Annu. Rev. Environ. Resour. 28, 429–461.
Morley, B., 2012. Empirical evidence on the effectiveness of environmental taxes. Appl. Econ. Lett. 19 (18), 1817–1820.
Murdoch, J.C., Sandler, T., 1997. The voluntary provision of a pure public good: the case of reduced CFC emissions and the Montreal Protocol. J. Public Econ. 63, 331–349.
Nordhaus, W., 2015. Climate clubs: overcoming free-riding in international climate policy. Am. Econ. Rev. 105 (4), 1339–1370.
Ringquist, E.J., Kostadinova, T., 2005. Assessing the effectiveness of international environmental agreements: the case of the 1985 Helsinki protocol. Am. J. Pol. Sci. 49 (1), 86–102.
Scott, M.E., 2007. Setting, and evaluating the effectiveness of, environmental policy. Environmetrics 18, 333–343.
Sunstein, C.R., 2007. Of Montreal and Kyoto: a tale of two protocols. Harv. Environ. Law Rev. 31, 1–65.
United Nations, 1998. Kyoto Protocol to the United Nations Framework Kyoto Protocol to the United Nations Framework. Technical report, United Nations, Kyoto.
Vollenweider, J., 2013. The effectiveness of international environmental agreements. Int. Environ. Agreements Polit. Law Econ. 13 (3), 343–367.
Wiener, J.B., 2007. Think globally, act globally: the limits of local climate policies. Univ. Penn. Law Rev. 155 (6), 1961–1979.
Xu, Y., 2017. Generalized synthetic control method: causal inference with interactive fixed effects models. Polit. Anal. 25 (1), 57–76.